THE JOURNAL OF PHILOSOPHY

VOLUME LXX, NO. 19, NOVEMBER 8, 1973

MATHEMATICAL TRUTH *

ALTHOUGH this symposium is entitled "Mathematical Truth," I will also discuss issues which are somewhat broader but which nevertheless have the notion of mathematical truth at their core, which themselves depend on how truth in mathematics is properly explained. The most important of these is mathematical knowledge. It is my contention that two quite distinct kinds of concerns have separately motivated accounts of the nature of mathematical truth: (1) the concern for having a homogeneous semantical theory in which semantics for the propositions of mathematics parallel the semantics for the rest of the language,[1] and (2) the concern that the account of mathematical truth mesh with a reasonable epistemology. It will be my general thesis that almost all accounts of the concept of mathematical truth can be identified with serving one or another of these masters *at the expense of the other*. Since I believe further that both concerns must be met by any adequate account, I find myself deeply

* To be presented at a symposium on Mathematical Truth, sponsored jointly by the American Philosophical Association, Eastern Division, and the Association for Symbolic Logic, December 27, 1973.

Commentators will be Oswaldo Chateaubriand and Saul Kripke; their comments are not available at this time. Various segments of an early (1967) version of this paper have been read at Berkeley, Harvard, Chicago Circle, Johns Hopkins, New York University, Princeton, and Yale. I am grateful for the help I received on these occasions, as well as for many comments from my colleagues at Princeton, both students and faculty. I am particularly indebted to Dick Grandy, Hartry Field, Adam Morton, and Mark Steiner. That these have not resulted in more significant improvements is due entirely to my own stubbornness. The present version is an attempt to summarize the essentials of the longer paper while making minor improvements along the way. The original version was written during 1967/68 with the generous support of the John Simon Guggenheim Foundation and Princeton University. This is gratefully acknowledged.

1 I am indulging here in the fiction that we *have* semantics for "the rest of language," or, more precisely, that the proponents of the views that take their impetus from this concern often think of themselves as having such semantics, at least for philosophically important segments of the language.

dissatisfied with any package of semantics and epistemology that purports to account for truth and knowledge both within and outside of mathematics. For, as I will suggest, accounts of truth that treat mathematical and nonmathematical discourse in relevantly similar ways do so at the cost of leaving it unintelligible how we can have any mathematical knowledge whatsoever; whereas those which attribute to mathematical propositions the kinds of truth conditions we can clearly know to obtain, do so at the expense of failing to connect these conditions with any analysis of the sentences which shows how the assigned conditions are conditions of their *truth*. What this means must ultimately be spelled out in some detail if I am to make out my case, and I cannot hope to do that within this limited context. But I will try to make it sufficiently clear to permit you to judge whether or not there is likely to be anything in the claim.

I take it to be obvious that any philosophically satisfactory account of truth, reference, meaning, and knowledge must embrace them all and must be adequate for all the propositions to which these concepts apply.[2] An account of knowledge that *seems* to work for certain empirical propositions about medium-sized physical objects but which fails to account for more theoretical knowledge is unsatisfactory—not only because it is incomplete, but because it may be incorrect as well, even as an account of the things it seems to cover quite adequately. To think otherwise would be, among other things, to ignore the interdependence of our knowledge in different areas. And similarly for accounts of truth and reference. A theory of truth for the language we speak, argue in, theorize in, mathematize in, etc., should by the same token provide similar truth conditions for similar sentences. The truth conditions assigned to two sentences containing quantifiers should reflect in relevantly similar ways the contribution made by the quantifiers. Any departure from a theory thus homogeneous would have to be strongly motivated to be worth considering. Such a departure, for

[2] I shall in fact have nothing to say about meaning in this paper. I believe that the concept is in much deserved disrepute, but I don't dismiss it for all that. Recent work, most notably by Kripke, suggests that what passed for a long time for meaning—namely the Fregean "sense"—has less to do with truth than Frege or his immediate followers thought it had. Reference is what is presumably most closely connected with truth, and it is for *this* reason that I will limit my attention to reference. If it is granted that change of reference can take place without a corresponding change in meaning, and that truth is a matter of reference, then talk of meaning is largely beside the point of the cluster of problems that concern us in this paper. These comments are not meant as arguments, but only as explanation.

example, might manifest itself in a theory that gave an account of the contribution of quantifiers in mathematical reasoning different from that in normal everyday reasoning about pencils, elephants, and vice-presidents. David Hilbert urged such an account in "On the Infinite," [3] which is discussed briefly below. Later on, I will try to say more about what conditions I would expect a satisfactory general theory of truth for our language to meet, as well as more about how such an account is to mesh with what I take to be a reasonable account of knowledge. Suffice it to say here that, although it will often be convenient to present my discussion in terms of theories of mathematical truth, we should always bear in mind that what is really at issue is our over-all philosophical view. I will argue that, *as an over-all view,* it is unsatisfactory—not so much because we lack a seemingly satisfactory account of mathematical truth or because we lack a seemingly satisfactory account of mathematical knowledge—as because we lack any account that satisfactorily brings the two together. I hope that it is possible ultimately to produce such an account; I hope further that this paper will help to bring one about by bringing into sharper focus some of the obstacles that stand in its way.

I. TWO KINDS OF ACCOUNT

Consider the following two sentences:

(1) There are at least three large cities older than New York.

(2) There are at least three perfect numbers greater than 17.

Do they have the same logicogrammatical form? More specifically, are they both of the form

(3) There are at least three *FG*'s that bear *R* to *a*.

where 'There are at least three' is a numerical quantifier eliminable in the usual way in favor of existential quantifiers, variables, and identity; '*F*' and '*G*' are to be replaced by one-place predicates, '*R*' by a two-place predicate, and '*a*' by the name of an element of the universe of discourse of the quantifiers? What are the truth conditions of (1) and (2)? Are they relevantly parallel? Let us ignore both the vagueness of 'large' and 'older than' and the peculiarities of attributive-adjective constructions in English which make a large city not something large and a city but more (although not exactly) like something large *for* a city. With those complications set aside, it seems clear that (3) accurately reflects the form of (1) and thus that (1) will be true if and only if the thing named by the expression replacing '*a*' ('New York') bears the relation designated by the ex-

[3] Translated and reprinted in Paul Benacerraf and Hilary Putnam, eds., *Philosophy of Mathematics* (Englewood Cliffs, N.J.: Prentice-Hall, 1964).

pression replacing '*R*' ('① is older than ②') to at least three elements (of the domain of discourse of the quantifiers) which satisfy the predicates replacing '*F*' and '*G*' ('large' and 'city', respectively). This, I gather, is what a suitable truth definition would tell us. And I think it's right. Thus, if (1) is true, it is because certain cities stand in a certain relation to each other, etc.

But what of (2)? May we use (3) in the same way as a matrix in spelling out the conditions of *its* truth? That sounds like a silly question to which the obvious answer is "Of course." Yet the history of the subject (the philosophy of mathematics) has seen many other answers. Some (including one of my past and present selves[4]), reluctant to face the consequences of combining what I shall dub such a "standard" semantical account with a platonistic view of the nature of numbers, have shied away from supposing that numerals are names and thus, by implication, that (2) is of the form (3). David Hilbert (*op. cit.*) chose a different but equally divergent approach, in his case in an attempt to arrive at a satisfactory account of the use of the notion of infinity in mathematics. On one construal, Hilbert can be seen as segregating a class of statements and methods, those of "intuitive" mathematics, as those which needed no further justification. Let us suppose that these are all "finitely verifiable" in some sense that is not precisely specified. Statements of arithmetic that do not share this property—typically, certain statements containing quantifiers—are seen by Hilbert as instrumental devices for going from "real" or "finitely verifiable" statements to "real" statements, much as an instrumentalist regards theories in natural science as a way of going from observation sentences to observation sentences. These mathematically "theoretical" statements Hilbert called "ideal elements," likening their introduction to the introduction of points "at infinity" in projective geometry: they are introduced as a convenience to make simpler and more elegant the theory of the things you really care about. If their introduction does not lead to contradiction and if they have these other uses, then it is justified: hence the search for a consistency proof for the full system of first-order arithmetic.

If this is a reasonable, if sketchy, account of Hilbert's view, it indicates that he did not regard all quantified statements semantically on a par with one another. A semantics for arithmetic as he viewed it would be very hard to give. But hard or not, it would certainly not treat the quantifier in (2) in the same way as the

[4] See my "What Numbers Could Not Be," *Philosophical Review*, LXXIV, 1 (January 1965): 47–73.

quantifier in (1). Hilbert's view as outlined represents a flat denial that (3) is the model according to which (2) is constructed.

On other such accounts, the truth conditions for arithmetic sentences are given as their formal derivability from specified sets of axioms. When coupled with the desire to attribute a truth value to each closed sentence of arithmetic, these views were torpedoed by the incompleteness theorems. They could be restored at least to internal consistency either by the liberalization of what counts as derivability (e.g., by including the application of an ω-rule in permissible derivations) or by abandoning the desire for completeness. For lack of a better term and because they almost invariably key on the syntactic (combinatorial) features of sentences, I will call such views "combinatorial" views of the determinants of mathematical truth. The leading idea of combinatorial views is that of assigning truth values to arithmetic sentences on the basis of certain (usually proof-theoretic) syntactic facts about them. Often, truth is defined as (formal) derivability from certain axioms. (Frequently a more modest claim is made—the claim to truth-in-*S*, where *S* is the particular system in question.) In any event, in such cases truth is conspicuously not explained in terms of reference, denotation, or satisfaction. The "truth" predicate is syntactically defined.

Similarly, certain views of truth in arithmetic on which the Peano axioms are claimed to be "analytic" of the concept of number are also "combinatorial" in my sense. And so are conventionalist accounts, since what marks them as conventionalist is the contrast between them and the "realist" account that analyzes (2) by assimilating it to (1), via (3).

Finally, to make one further distinction, a view is not automatically "combinatorial" if it interprets mathematical propositions as being about combinatorial matters, either self-referentially or otherwise. For such a view might analyze mathematical propositions in a "standard" way in terms of the names and quantifiers they might contain and in terms of the properties they ascribe to the objects within their domains of discourse—which is to say that the underlying concept of *truth* is essentially Tarski's. The difference is that its proponents, although realists in their analysis of mathematical language, part ways with the platonists by construing the mathematical universe as consisting exclusively of mathematically unorthodox objects: Mathematics for them is limited to metamathematics, and that to syntax.

I will defer to later sections my assessment of the relative merits of these various approaches to the truth of such sentences as (2). At this point I wish only to introduce the distinction between, on the one hand, those views which attribute the obvious syntax (and the obvious semantics) to mathematical statements, and, on the other, those which, ignoring the apparent syntax and semantics, attempt to state truth conditions (or to specify and account for the existing distribution of truth values) on the basis of what are evidently non-semantic syntactic considerations. Ultimately I will argue that each kind of account has its merits and defects: each addresses itself to an important component of a coherent over-all philosophic account of truth and knowledge.

But what are these components, and how do they relate to one another?

II. TWO CONDITIONS

A. The first component of such an over-all view is more directly concerned with the concept of truth. For present purposes we can state it as the requirement that there be an over-all theory of truth in terms of which it can be certified that the account of mathematical truth is indeed an account of mathematical *truth*. The account should imply truth conditions for mathematical propositions that are evidently conditions of their truth (and not *simply*, say, of their theoremhood in some formal system). This is not to *deny* that being a theorem of some system can be a truth condition for a given proposition or class of propositions. It is rather to require that any theory that proffers theoremhood as a condition of truth also *explain the connection between truth and theoremhood.*

Another way of putting this first requirement is to demand that any theory of mathematical truth be in conformity with a general theory of truth—a theory of truth theories, if you like—which certifies that the property of sentences that the account calls "truth" is indeed truth. This, it seems to me, can be done only on the basis of some general theory for at least the language as a whole (I assume that we skirt paradoxes in some suitable fashion). Perhaps the applicability of this requirement to the present case amounts only to a plea that the semantical apparatus of mathematics be seen as part and parcel of that of the natural language in which it is done, and thus that whatever *semantical* account we are inclined to give of names or, more generally, of singular terms, predicates, and quantifiers in the mother tongue include those parts of the mother tongue which we classify as mathematese.

I suggest that, if we are to meet this requirement, we shouldn't be satisfied with an account that fails to treat (1) and (2) in parallel fashion, on the model of (3). There may well be *differences,* but I expect these to emerge at the level of the analysis of the reference of the singular terms and predicates. I take it that we have only one such account: Tarski's, and that its essential feature is to define truth in terms of reference (or satisfaction) on the basis of a particular kind of syntactico-semantical analysis of the language, and thus that any putative analysis of mathematical truth must be an analysis of a concept which is a truth concept at least in Tarski's sense. Suitably elaborated, I believe this requirement to be inconsistent with all the accounts that I have termed "combinatorial." On the other hand, the account that assimilates (2) above to (1) and (3) obviously meets this condition, as do many variants of it.

B. My second condition on an over-all view presupposes that we have mathematical knowledge and that such knowledge is no less knowledge for being mathematical. Since our knowledge is of truths, or can be so construed, an account of mathematical truth, to be acceptable, must be consistent with the possibility of having mathematical knowledge: the conditions of the truth of mathematical propositions cannot make it impossible for us to know that they are satisfied. This is not to argue that there cannot be unknowable truths—only that not all truths can be unknowable, for we know some. The minimal requirement, then, is that a satisfactory account of mathematical truth must be consistent with the possibility that some such truths be knowable. To put it more strongly, the concept of mathematical truth, as explicated, must fit into an over-all account of knowledge in a way that makes it intelligible how we have the mathematical knowledge that we have. An acceptable semantics for mathematics must fit an acceptable epistemology. For example, if I know that Cleveland is between New York and Chicago, it is because there exists a certain relation between the truth conditions for that statement and my present "subjective" state of belief (whatever may be our accounts of truth and knowledge, they must connect with each other in this way). Similarly, in mathematics, it must be possible to link up what it is for p to be true with my belief that p. Though this is extremely vague, I think one can see how the second condition tends to rule out accounts that satisfy the first, and to admit many of those which do not. For a typical "standard" account (at least in the case of number theory or set theory) will depict truth conditions in terms of conditions on objects whose nature, as normally conceived, places them beyond the reach of the better understood means of human cognition (e.g., sense perception and the like). The "combinatorial" accounts, on the other hand, usually arise from a sensitivity to precisely this fact and are hence almost always motivated by epistemological concerns. Their virtue lies in providing an account of mathematical propositions based on the procedures we follow in justifying truth claims in mathematics: namely, proof. It is not surprising that *modulo* such accounts of mathematical truth, there is little mystery about how we can obtain mathematical knowledge. We need only account for our ability to produce and survey formal proofs.[5] However, squeezing the balloon at that point apparently makes it bulge on the side of truth: the more nicely we tie up the concept of proof, the more closely we link the definition of proof to combinatorial (rather than semantic) features, the more difficult it is to connect it up with the truth of what is being thus "proved"—or so it would appear.

These then are the two requirements. Separately, they seem innocuous enough. In the balance of this paper I will both defend them further and flesh out the argument that jointly they seem to rule out almost every account of mathematical truth that has been proposed. I will consider in turn the two basic approaches to mathematical truth that I mentioned above, weighing their relative advantages in light of the two fundamental principles that I am advancing. I hope that the principles themselves will receive some illumination and support as I do so.

III. THE STANDARD VIEW

I call the "platonistic" account that analyzes (2) as being of the form (3) "the standard view." Its virtues are many, and it is worth enumerating them in some detail before passing to a consideration of its defects.

As I have already pointed out, this account assimilates the logical form of mathematical propositions to that of apparently similar empirical ones: empirical and mathematical propositions alike contain predicates, singular terms, quantifiers, etc.

But what of sentences that are not composed (or correctly analyzable as being composed of) names, predicates, and quantifiers? More directly to the point, what of sentences that do not belong to the kind of language for which Tarski has showed us how to define truth? I would say that we need for such languages (if there are any) an account of truth of the sort that Tarski supplied for

[5] Properly done, this is of course an enormous task. Nevertheless it sets to one side accounting for the burden that is borne by the semantics of the system and by our understanding of it, concentrating instead on our ability to determine that certain formal objects have certain syntactically defined properties.

"referential" languages. I assume that the truth conditions for the language (e.g., English) to which mathematese appears to belong are to be elaborated much along the lines that Tarski articulated. So, to some extent, the question posed in the previous section—How are truth conditions for (2) to be explained?—may be interpreted as asking whether the sublanguage of English in which mathematics is done is to receive the same sort of analysis as I am assuming is appropriate for much of the rest of English. If so, then the qualms I shall sketch in the next section concerning how to fit mathematical knowledge into an over-all epistemology clearly apply—though they can perhaps be laid to rest by a suitable modification of theory. If, on the other hand, mathematese is not to be analyzed along referential lines, then we are clearly in need not only of an account of truth (i.e., a semantics) for this new kind of language, but also of a new *theory of truth theories* that relates truth for referential (quantificational) languages to truth for these new (newly analyzed) languages. Given such an account, the task of accounting for mathematical knowledge would still remain; but it would presumably be an easier task, since the new semantical picture of mathematese would in most cases have been prompted by epistemological considerations. However, I do not give this alternative serious consideration in this paper because I don't think that anyone has ever actually chosen it. For to choose it is explicitly to consider *and reject* the "standard" interpretation of mathematical language, despite its superficial and initial plausibility, and then to provide an alternative semantics as a substitute.[6] The "combinatorial" theorists whom I discuss or refer to have usually wanted to have their cake and eat it too: they have not realized that the truth conditions that their account supplies for mathematical language have not been connected to the referential semantics which they assume is *also* appropriate for that language. Perhaps the closest candidate for an exception is Hilbert in the view I sketched briefly in the opening pages of this paper. But to pursue this further here would take us too far afield. Let us return, therefore, to our praise of the "standard view."

One of its primary advantages is that the truth definitions for individual mathematical theories thus construed will have the same recursion clauses as those employed for their less lofty empirical cousins. Or to put it another way, they can all be taken as parts of

[6] I sometimes think this is one of the things that Hilary Putnam wants to do in his stimulating article "Mathematics without Foundations," this JOURNAL, LXIV, 1 (Jan. 19, 1967): 5–22.

the same language for which we provide a single account for quantifiers regardless of the subdiscipline under consideration. Mathematical and empirical disciplines will not be distinguished in point of logical grammar. I have already underscored the importance of this advantage: it means that the logicogrammatical theory we employ in less recondite and more tractable domains will serve us well here. We can do with one, uniform, account and need not invent another for mathematics. This should hold true on virtually any grammatical theory coupled with semantics adequate to account for truth. My bias for what I call a Tarskian theory stems simply from the fact that he has given us the only viable systematic general account we have of truth. So, one consequence of the economy attending the standard view is that logical relations are subject to uniform treatment: they are invariant with subject matter. Indeed, they help define the concept of "subject matter." The same rules of inference may be used and their use accounted for by the same theory which provides us with our ordinary account of inference, thus avoiding a double standard. If we reject the standard view, mathematical inference will need a new and special account. As it is, standard uses of quantifier inferences are justified by some sort of soundness proof. The formalization of theories in first-order logic requires for *its* justification the assurance (provided by the Completeness theorem) that all the logical consequences of the postulates will be forthcoming as theorems. The standard account delivers these guarantees. The obvious answers seem to work. To reject the standard view is to discard these answers. New ones would have to be found.

So much for the obvious virtues of this account. What are its faults?

As I suggested above, the principal defect of the standard account is that it appears to violate the requirement that our account of mathematical truth be susceptible to integration into our over-all account of knowledge. Quite obviously, to make out a persuasive case to this effect it would be necessary to sketch the epistemology I take to be at least roughly correct and on the basis of which mathematical truths, standardly construed, do not seem to constitute knowledge. This would require a lengthy detour through the general problems of epistemology. I will leave that to another time and content myself here with presenting a brief summary of the salient features of that view which bear most immediately on our problem.

IV. KNOWLEDGE

I favor a causal account of knowledge on which for *X* to know that *S* is true requires some causal relation to obtain between *X* and the referents of the names, predicates, and quantifiers of *S*. I believe in addition in a causal theory of *reference*, thus making the link to my saying knowingly that *S doubly* causal. I hope that what follows will dispel some of the fog which surrounds this formulation.

For Hermione to know that the black object she is holding is a truffle is for her (or at least requires her) to be in a certain (perhaps psychological) state.[7] It also requires the cooperation of the rest of the world, at least to the extent of permitting the object she is holding to be a truffle. Further—and this is the part I would emphasize—in the normal case, that the black object she is holding is a truffle must figure in a suitable way in a causal explanation of her belief that the black object she is holding is a truffle. But what is a "suitable way"? I will not try to say. A number of authors have published views that seem to point in this direction,[8] and, despite differences among them, there seems to be a core intuition which they share and which I think is correct although very difficult to pin down.

That some such view must be correct and underlies our conception of knowledge is indicated by what we would say under the following circumstances. It is claimed that *X* knows that *p*. We think that *X* could not know that *p*. What reasons can we offer in support of our view? If we are satisfied that *X* has normal inferential powers, that *p* is indeed true, etc., we are often thrown back on arguing that *X* could not have come into possession of the relevant evidence or reasons: that *X*'s four-dimensional space-time worm does not make the necessary (causal) contact with the grounds of the truth of the proposition for *X* to be in possession of evidence ade-

[7] If possible, I would like to avoid taking any stand on the cluster of issues in the philosophy of mind or psychology concerning the nature of psychological states. Any view on which Hermione can learn that the cat is on the mat by looking at a real cat on a real mat will do for my purposes. If looking at a cat on a mat puts Hermione into a state and you wish to call that state a physical, or psychological, or even physiological state, I will not object so long as it is understood that such a state, if it is her state of knowledge, is causally related in an appropriate way to the cat's having been on the mat when she looked. If there is no such state, then so much the worse for my view.

[8] To cite but a few: Gilbert H. Harman, *Thought* (Princeton, N.J.: University Press, 1973); Alvin I. Goldman, "A Causal Theory of Knowing," this JOURNAL, LXIV, 12 (June 22, 1967): 357–372; Brian Skyrms, "The Explication of '*X* knows that *p*'," *ibid.*, 373–389.

quate to support the inference (if an inference was relevant). The proposition *p* places restrictions on what the world can be like. Our knowledge of the world, combined with our understanding of the restrictions placed by *p,* given by the truth conditions of *p,* will often tell us that a given individual could not have come into possession of evidence sufficient to come to know *p,* and we will thus deny his claim to the knowledge.

As an account of our knowledge about medium-sized objects, in the present, this is along the right lines. It will involve, causally, some direct reference to the facts known, and, through that, reference to these objects themselves. Furthermore, such knowledge (of houses, trees, truffles, dogs, and bread boxes) presents the clearest case and the easiest to deal with.

Other cases of knowledge can be explained as being based on inferences based on cases such as these, although there must evidently be interdependencies. This is meant to include our knowledge of general laws and theories, and, through them, our knowledge of the future and much of the past. This account follows closely the lines that have been proposed by empiricists, but with the crucial modification introduced by the explicitly causal condition mentioned above—but often left out of modern accounts, largely because of attempts to draw a careful distinction between "discovery" and "justification."

In brief, in conjunction with our other knowledge, we use *p* to determine the range of possible relevant evidence. We use what we know of *X* (the putative knower) to determine whether there could have been an appropriate kind of interaction, whether X's current belief that *p* is causally related in a suitable way with what is the case because *p* is true—whether his evidence is drawn from the range determined by *p*. If not, then *X* could not know that *p*. The connection between what must be the case if *p* is true and the causes of *X*'s belief can vary widely. But there is always *some* connection, and the connection relates the grounds of *X*'s belief to the subject matter of *p*.

It must be possible to establish an appropriate sort of connection between the truth conditions of *p* (as given by an adequate truth definition for the language in which *p* is expressed) and the grounds on which *p* is said to be known, at least for propositions that one must *come to know*—that are not innate. In the absence of this, no connection has been established between *having those grounds* and *believing a proposition which is true.* Having those grounds cannot be fitted into an explanation of *knowing p*. The link between *p* and justifying a belief in *p on those grounds* cannot be made. But for that knowledge which is properly regarded as some form of justified true belief, then the link *must* be made. (Of course not *all* knowledge need be justified true belief for the point to be a sound one.)

It will come as no surprise that this has been a preamble to pointing out that combining *this* view of knowledge with the "standard" view of mathematical truth makes it difficult to see how mathematical knowledge is possible. If, for example, numbers are the kinds of entities they are normally taken to be, then the connection between the truth conditions for the statements of number theory and any relevant events connected with the people who are supposed to have mathematical knowledge cannot be made out.[9] It will be impossible to account for how anyone knows any properly number-theoretical propositions. This second condition on an account of mathematical truth will not be satisfied, because we have no account of how we know that the truth conditions for mathematical propositions obtain. One obvious answer—that some of these propositions are true if and only if they are derivable from certain axioms via certain rules—will not help here. For, to be sure, we can ascertain that *those* conditions obtain. But in such a case, what we lack is the link between truth and proof, when truth is directly defined in the standard way. In short, although it may be a truth condition of certain number-theoretic propositions that they be derivable from certain axioms according to certain rules, *that* this is a truth condition must also follow from the account of *truth* if the condition referred to is to help connect truth and knowledge, if it is by their proofs that we know mathematical truths.

Of course, given some set-theoretical account of arithmetic, both the syntax and the semantics of *arithmetic* can be set out so as superficially to meet the conditions we have laid down. But the regress that this invites is transparent, for the same questions must then be asked about the set theory in terms of which the answers are couched.

V. TWO EXAMPLES

There are many accounts of mathematical truth and mathematical knowledge. The theses I have been defending are intended to apply to them all. Rather than try to be comprehensive, however, I will devote these last few pages to the examination of two representa-

[9] For an expression of healthy skepticism concerning this and related points, see Mark Steiner, "Platonism and the Causal Theory of Knowledge," this JOURNAL, LXX, 3 (Feb. 8, 1973): 57–66.

tive cases: one "standard" view and one "combinatorial" view. First the standard account, as expressed by one of its most explicit and lucid proponents, Kurt Gödel.

Gödel is thoroughly aware that on a realist (i.e., standard) account of mathematical truth our explanation of how we know the basic postulates must be suitably connected with how we interpret the referential apparatus of the theory. Thus, in discussing how we can resolve the continuum problem, once it has been shown to be undecidable by the accepted axioms, he paints the following picture:

> . . . the objects of transfinite set theory . . . clearly do not belong to the physical world and even their indirect connection with physical experience is very loose . . .
>
> But, despite their remoteness from sense experience, we do have a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception, which induces us to build up physical theories and to expect that future sense perceptions will agree with them and, moreover, to believe that a question not decidable now has meaning and may be decided in the future.[10]

I find this picture both encouraging and troubling. What troubles me is that without an account of *how* the axioms "force themselves upon us as being true," the analogy with sense perception and physical science is without much content. For what is missing is *precisely* what my second principle demands: an account of the link between our cognitive faculties and the objects known. In physical science we have at least a start on such an account, and it is causal. We accept as knowledge only those beliefs which we can appropriately relate to our cognitive faculties. Quite appropriately, our conception of knowledge goes hand in hand with our conception of ourselves as knowers. To be sure, there is a *superficial* analogy. For, as Gödel points out, we "verify" axioms by deducing consequences from them concerning areas in which we seem to have more direct "perception" (clearer intuitions). But we are never told how we know even these, clearer, propositions. For example, the "verifiable" consequences of axioms of higher infinity are (otherwise undecidable) number-theoretical propositions which themselves are "verifiable" by computation up to any given integer. But the story, to be helpful anywhere, must tell us how we know

[10] "What Is Cantor's Continuum Problem?" revised version in Benacerraf and Putnam, *op. cit.*, p. 271.

statements of computational arithmetic—*if they mean what the standard account would have them mean.* And *that* we are not told. So the analogy is at best superficial.

So much for the troubling aspects. More important perhaps and what I find encouraging is the evident basic agreement which motivates Gödel's attempt to draw a parallel between mathematics and empirical science. He sees, I think, that something must be said to bridge the chasm, created by his realistic and platonistic interpretation of mathematical propositions, between the entities that form the subject matter of mathematics and the human knower. Insead of tinkering with the logical form of mathematical propositions or with the nature of the objects known, he postulates a special faculty through which we "interact" with these objects. We seem to agree on the analysis of the fundamental problem, but clearly disagree about the epistemological issue—about what avenues are open to us through which we may come to know things.

If our account of empirical knowledge is acceptable, it must be in part because it tries to make the connection evident in the case of our theoretical knowledge, where it is not *prima facie* clear how the causal account is to be filled in. Thus, when we come to mathematics, the absence of a coherent account of how our mathematical intuition is connected with the truth of mathematical propositions renders the over-all account unsatisfactory.

To introduce a speculative historical note, with some foundation in the texts, it might not be unreasonable to suppose that Plato had recourse to the concept of *anamnesis* at least in part to explain how, given the nature of the forms as he depicted them, one could ever have knowledge of them.[11]

The "combinatorial" view of mathematical truth has epistemological roots. It starts from the proposition that, whatever may be the "objects" of mathematics, our knowledge is obtained from proofs. Proofs are or can be (for some, must be) written down or spoken; mathematicians can survey them and come to agree that they *are* proofs. It is largely through these proofs that mathematical knowledge is obtained and transmitted. In short, this aspect of mathematical knowledge—its (essentially linguistic) means of production and transmission gives their impetus to the class of views that I call "combinatorial."

Noticing the role of proofs in the production of knowledge, it

[11] "The soul, then, as being immortal, and having been born again many times, and having seen all things that exist, whether in this world or in the world below, has knowledge of them all" (Plato, *Meno,* 81).

seeks the grounds of truth in the proofs themselves. Combinatorial views receive additional impetus from the realization that the platonist casts a shroud of mystery over how knowledge can be obtained at all. Add that realization to the belief that mathematics is a child of our own begetting (mathematical discovery, on these views, is seldom discovery about an independent reality), and it is not surprising that one looks for acts of conception to account for the birth. Many accounts of mathematical truth fall under this rubric. Perhaps almost all. I have mentioned several in passing, and I discussed Hilbert's view in "On the Infinite" very briefly. The final example I wish to consider is that of conventionalist accounts —the cluster of views that the truths of logic and mathematics are true (or can be made true) in virtue of explicit conventions where the conventions in question are usually the postulates of the theory. Once more, I will probably do them all an injustice by lumping together a number of views which their proponents would most certainly like to keep apart.

Quine, in his classic paper on this subject,[12] has dealt clearly, convincingly, and decisively with the view that the truths of *logic* are to be accounted for as the products of convention—far better than I could hope to do here. He pointed out that, since we must account for infinitely many truths, the characterization of the eligible sentences as truths must be wholesale rather than retail. But wholesale characterization can proceed only via general principles—and, if we are supposed not to understand any logic at all, we cannot extract the individual instances from the general principles: we would need logic for such a task.

Persuasive as this may be, I wish to add another argument—not because I think this dead horse needs further flogging, but both because Quine's argument is limited to the case of logic and because the principal points I wish to bring out do not emerge sufficiently from it. Indeed, Quine grants the conventionalist certain principles I should like to deny him. In resting his case against conventionalism on the need for a wholesale characterization of infinitely many truths, Quine concedes that were there only finitely many truths to be reckoned with, the conventionalist might have a chance to make out his case. He says:

> If truth assignments could be made one by one, rather than an infinite number at a time, the above difficulty would disappear; truths of logic . . . would simply be asserted severally by fiat, and the problem

[12] W. V. Quine, "Truth by Convention," reprinted in Benacerraf and Putnam, *op. cit.*

of inferring them from more general conventions would not arise. (p. 344).

Thus, if some way could be found to make sentences of logic wear their truth values upon their sleeves, the objections to the conventionalist account of truth would disappear—for we would have determined truth values for all the sentences, which is all that one could ask.

I wonder, however, what such a sprinkling of the word 'true' would accomplish. Surely it cannot suffice in order to determine a concept of truth to assign values to each and every sentence of the language [suppose now that the language is set theory, in some first-order formalization] (let those with an even number of horseshoes be "true").

What would make such an assignment of the predicate 'true' the determination of *the concept of truth?* Simply the use of that monosyllable? Tarski has suggested that satisfaction of Convention T is a necessary and sufficient condition on a definition of truth for a particular language.[13] A mere (recursive) distribution of truth values can be parlayed into a truth theory that satisfies convention T. We can rest with that provided we are prepared to beg what I think is the main question and ignore the concept of translation that occurs in its (Convention T's) formulation. What would be missing, hard as it is to state, is the theoretical apparatus employed by Tarski in providing truth definitions, i.e., the analysis of truth in terms of the "referential" concepts of naming, predication, satisfaction, and quantification. A definition that does not proceed by the customary recursion clauses for the customary grammatical forms may not be adequate, even if it satisfies Convention T. The explanation must proceed through reference and satisfaction and, furthermore, must be supplemented with an account of reference itself. But the defense of this last claim is too involved a matter to take up here.[14]

[13] Alfred Tarski, "The Concept of Truth for Formalized Languages," reprinted in Tarski, *Logic, Semantics, and Metamathematics* (New York: Oxford, 1956). Convention T is stated on pp. 187/8 as follows:

CONVENTION T. A formally correct definition of the symbol 'Tr', formulated in the metalanguage, will be called *an adequate definition of truth* if it has the following consequences:

(α) all sentences which are obtained from the expression '$x \epsilon \mathrm{Tr}$ if and only if p' by substituting for the symbol 'x' a structural-descriptive name of any sentence of the language in question and for the symbol 'p' the expression which forms the translation of this sentence into the metalanguage;

(β) the sentence 'for any x, if $x \epsilon \mathrm{Tr}$ then $x \epsilon \mathrm{S}$' (in other words '$\mathrm{Tr} \subseteq \mathrm{S}$').

[14] For an excellent presentation of a similar view, see Hartry Field, "Tarski's Theory of Truth," this JOURNAL, LXIX, 13 (July 13, 1972): 347–375.

The Quine of "Truth by Convention" felt that to determine the truth values of all the contexts that contain a word suffices to determine its reference. That *might* be so, if we already had the concept of truth and chased the reference of the term that interested us down through the truth definition. But there seems to be something patently wrong with trying to fix the concept of truth *itself* in this way. In so doing, we throw away the very crutch which enables that method to work for other concepts. Truth and reference go hand in hand. Our concept of truth, insofar as we have one, proceeds through the mediation of the concepts Tarski has used to define it for the class of languages he has considered—the essence of Tarski's contribution goes much further than Convention T, but includes the schemata for the actual definition as well: an analysis of truth for a language that did not proceed through the familiar devices of predication, quantification, etc., should not give us satisfaction.

If this is at all near the mark, then it should be clear why "combinatorial" views of the nature of mathematical truth fail on my account. They avoid what seems to me to be the necessary route to an account of truth: through the subject matter of the propositions whose truth is being defined. Motivated by epistemological considerations, they come up with truth conditions whose satisfaction or nonsatisfaction mere mortals can ascertain; but the price they pay is their inability to connect these so-called "truth conditions" with the truth of the propositions for which they are conditions.

Even if it is granted that the truths of first-order logic do not stem from conventions, it might still be claimed that the rest of mathematics (set theory, for logicists; set theory, number theory, and other things for nonlogicists) consists of conventions formalized in first-order logic. This view too is subject to the objection that such a concept of convention need not bring *truth* along with it.[15] Indeed it is clear that it does not. For, even ignoring more general objections, once the logic is fixed, it becomes possible that the conventions thus stipulated turn out to be inconsistent. Hence it cannot be maintained that setting down conventions *guarantees* truth. But if it does not *guarantee* truth, what distinguishes those

[15] Identical arguments will apply to the view, perhaps indistinguishable from this one, that the postulates constitute *implicit definitions* of existing concepts (as opposed to stipulating how new ones are to be understood), if that is advanced to explain how we know the axioms to be true (we learned the language by learning *these* postulates).

cases in which it provides for it from those in which it does not? Consistency cannot be the answer. To urge it as such is to *mis*construe the significance of the fact that *in*consistency is *proof* that truth has not been attained. The deeper reason once more is that postulational stipulation makes no connection between the propositions and their subject matter—stipulation does not provide for truth. At best, it limits the class of truth definitions (interpretations) consistent with the stipulations. But that is not enough.

To clarify the point, consider Russell's oft-cited dictum: "The method of 'postulating' what we want has many advantages; they are the same as the advantages of theft over honest toil." [16] On the view I am advancing, that's false. For with theft at least you come away with the loot, whereas implicit definition, conventional postulation, and their cousins are incapable of bringing truth. They are not only morally but practically deficient as well.

PAUL BENACERRAF

Princeton University

MATTER *

SCULPTORS are sometimes said to work in this or that material —marble or wood or terra cotta. The Greek sculptor Myron, according to ancient testimony, worked almost exclusively in bronze. But it could be misleading to put it this way. The fact is that Myron's statues were *made of* bronze: his famous Discobolus was a *bronze statue*. But it is unlikely that Myron did much actual work on, or with, the bronze of which the Discobolus was made. Indeed, it is unlikely that that bronze even existed at the time that Myron was doing his main work on the statue. The Discobolus itself has long since ceased to exist; no doubt some barbarian invader had it melted down and used its bronze to make a shield. But we know a good deal about it, owing to the descriptions of Lucian and Pliny, and to Roman copies of it, in marble, several of which have survived. Thus we know that it was hollow and that it was cast by the so-called "lost-wax" process. This means that Myron would have begun by modeling, somewhat roughly, a figure in clay. He would then have covered this clay model with a thin

[16] Bertrand Russell, *Introduction to Mathematical Philosophy* (London: Allen & Unwin, 1919), p. 71.

* To be presented in an APA symposium on Aristotle's Conception of Matter, December 27, 1973. Commentators will be John M. Cooper and Russell M. Dancy; see this JOURNAL, this issue, 696–698 and 698–699, respectively.