

On the Possibility of Constructing Truth-Conditions for Self-Referential Propositions

Patrick Colm Hogan
State University of New York at Buffalo

Despite the remarkable problems encountered by classificatory treatments self-referential propositions, virtually all 'solutions' proposed to the paradoxes generated by these propositions consist, as Andre Glucksmann has put it, "either in ruling out such propositions as nonsenses, absurdities, or in accepting them while making a hierarchical distinction."¹ Clearly, the ideal solution to paradoxes of self-reference would involve a) prevention of regeneration of the paradoxes at a level above that of the original occurrence, b) delimitation of clear criteria for the solution based upon arguments drawn from principles beyond those directly related to the paradoxes themselves (e.g., in the case of hierarchical solutions, a statement of and argument for conditions for hierarchical distinctions, beyond a mere *ad hoc* elimination of the paradoxical cases). Karl Popper has argued, quite cogently, that making a paradoxical self-reference meaningless merely regenerates the paradox at another level,² and Hans Regnell has pointed out that generally "the theory of the object-language and the meta-language does not state the necessary and sufficient conditions for the appearance of the antinomies. Nor does it make evident why these antinomies sometimes, but not always, appear when the distinction between object-language and meta-language is ignored."³ On the other hand, many solutions which seem to work require sacrifices (such as that of substitutivity of identicals) which most of us would be unwilling to make, as Fitch has noted.⁴

These problems, I believe, stem from asking the wrong sorts of questions about self-referential propositions--specifically, 'Might such propositions be classified as true or false?' rather than 'Under what conditions might such propositions really be true or false?' It seems to me just as odd to turn to formal logic to decide if 'This sentence is true' is true, as it would to decide if "The cat is on the mat" is true (or if "The cat is on the mat' is true" is true). Formal logic, Karl Popper has emphasized, deals with "the transmission of truth and the retransmission of falsity"⁵ from proposition to proposition; it does not

deal with the actual truth or falsity of propositions themselves. This is not to say, of course, that the delimitation of what 'objects' may be 'mapped onto T,F', the 'circumscription' of 'truth- and falsity-predicable objects' is irrelevant to formal logic, but merely that formal logic by itself and as formal cannot determine whether or not certain states of affairs obtain in the world--no amount of logical analysis will tell us if the cat is, in fact, on the mat, though it may tell us what else must be the case if the cat is on the mat, and what else must be the case if the cat is not so situated.' Thus formal logic does not, or should not, ask "Is 'This sentence is false' true or false?" but "If 'This sentence is false' is true, what follows, and if false, what follows?" The answer to this latter, hypothetical question, mistaken for an answer to the former, factual question, has given rise to the notion that assertions such as 'This sentence is false' are paradoxical. In fact, formal logic says nothing at all about any actual paradoxes resulting from the truth or falsity of such assertions, but only something about the paradoxes which would be generated if such assertions were either true or false.

To decide if an assertion is true or false we must begin by constructing the conditions under which such an assertion would be true, conditions defined by that assertion, and then proceed to see if these conditions obtain in the world. (We should never ask ourselves if an assertion is true without having first fully determined under what conditions it would be true.) Suppose, for example, that we construe the truth-conditions for "This sentence is false" and discover that they do (or do not) obtain, then the laws of logic will show us an actual (and not merely hypothetical) paradox. If, on the other hand, we find that these truth-conditions are not fully construable, as they are not fully defined, then no actual paradox can result, for the proposition cannot be true or false in the first place, there being no determinate truth-conditions which may be said to obtain in the world or not obtain. This may be clearer with another example. Let us suppose that we wish to determine the truth or falsity of 'He has a big nose', and let us suppose further that we have determined with context-appropriate accuracy what will constitute bigness with regard to noses. Let us suppose finally that the speaker (a psychotic?) literally has no one in mind when asserting 'He has a big nose' (and furthermore does not mean 'Someone has a big nose'). In this case the proposition does not fully define truth-conditions and is thus neither true nor false, as there are no determinate conditions which may be said to obtain or not obtain in the world and make that proposition true or false. It is my contention that 'undecidable' and 'paradoxical' self-referential propositions similarly do not fully

define truth-conditions, therefore can be neither true nor false, and therefore, cannot be paradoxical (i.e. both true and false) or undecidable (i.e. equally assertable as true or false and, therefore, again, both true and false). In other words, self-referential assertions lack a truth-value which might be transmitted or re-transmitted by the laws of logic and thereby generate paradoxes.

A necessary condition for a proposition to have a truth-value is, as we have seen, its having fully determinate truth-conditions, and this entails that all indexical elements in such propositions have definite referents. Given an assertion such as 'He has a big nose', the reference of 'he' must be definite (i.e. there must be such a reference) if the proposition is to be true or false, if it is to have truth-conditions which may or may not obtain in the world. When an assertion itself is the referent of an index and the metalinguistic assertion asserts the truth or falsity of the object-sentence, the same definiteness of reference must obtain in that object-sentence if the metalinguistic assertion is to be true. Let us take, for example, 'This sentence is true', where the sentence referred to is 'He has a big nose'; if the 'he' of the object sentence is indeterminate, then the metalinguistic assertion cannot be true; if on the other hand, it is determinate, then the metalinguistic assertion may be true or may not. In any case, the metalinguistic assertion itself has determinate truth-conditions which are unaffected by the indefiniteness of its object-sentence's truth-conditions--the former is true, simply, if 'He has a big nose' defines truth-conditions which obtain in the world. Similarly, if yet a third assertion were made concerning the truth of the first metalinguistic assertion (i.e. 'This sentence is true' referring to 'This sentence is true' referring to 'He has a big nose'), its truth-conditions would also be fully definite.

But let us take an assertion which is considered the same as its object sentence and, thus, having the same truth-conditions; clearly, if the truth-conditions of the latter are indeterminate, those of the former are indeterminate by substitutivity of identicals. Let us take, for example, 'This sentence is true', where the 'this sentence' is meant to refer to 'This sentence is true', which is identified with the metalinguistic 'This sentence is true' and which, thus, itself refers to 'This sentence is true', and so on. The first referent includes an indexical element, the referent of which includes an indexical element, the referent of which includes an indexical element, and so on infinitely. Thus, due to metalinguistic-sentence/object-sentence identity, each metalinguistic sentence is contingent for the definiteness of its truth-conditions upon the definiteness of the truth-conditions of its

object-sentence, but each object-sentence is itself a metalinguistic sentence, and thus is itself contingent in precisely the same fashion, due to the infinite generation of mutually identified indices. (An attempt to write out this sentence, in the form 'It is true that x', for $x = \text{it is true that } x$, would begin 'It is true that it is true that it is true that . . .' and continue thus indefinitely; in the form 'This sentence is true', such an attempt would begin with an infinite series of quotation marks.) Thus assertions of this form (i.e. self-referential assertions of a certain type) do not completely define truth-conditions, and are, therefore, neither true nor false, as their indexical elements are infinitely self-generating. (We should note here that such infinite self-generation cannot be avoided by circumlocutions of the form 'This sentence-token which I am now uttering is true' as sentence-tokens are not the sort of thing that can be true--they may be loud or short or garbled, but not true.)

Thus we conclude that 'This sentence does not have fully determinate truth-conditions' is true when the sentence referred to is the self-referential 'This sentence is true'. The truth-conditions for the first sentence are clearly determinate; it is true if the sentence to which it refers infinitely regenerates an indeterminate element in the definition of its truth-conditions. This is equally true if the sentence to which it refers is 'This sentence does not have fully determinate truth-conditions', if this is considered a separate sentence with different truth-conditions. On the other hand, if 'This sentence does not fully determine truth-conditions' is meant to refer to a sentence with which it is identified and, hence, to a sentence with identical truth-conditions, then the definition of the truth-conditions of the metalinguistic sentence is contingent upon the definiteness of the truth-conditions of the object-sentence, in the manner of the above examples, and, hence, its truth-conditions rest indefinite.

It is important to note that it does not follow from this argument that every self-referential proposition lacks fully definite truth-conditions, or, rather, that every self-referential assertion necessarily lacks definiteness in certain particulars claimed in that assertion true of the world. 'This sentence is being used as an example', for example, is fully definite in all particulars claimed true of the world and is, indeed, true. However, in assertions such as this, only a partial identification between metalinguistic sentence and object-sentence occurs; the metalinguistic assertion 'This sentence is being used as an example' is identified with its object-assertion only in its function as example, in the aspect relevant to the predication, and not in its truth-conditions. The truth-

conditions of the object-assertion are not even posited, one might say, in the metalinguistic assertion, for the truth-conditions of the former are entirely irrelevant to the truth-claims of the latter (i.e. the determination of the truth-conditions of a sentence does not even enter into the determination of whether or not that sentence is functioning as an example) and would not, therefore, be included in the truth-conditions of the latter, either as identified therewith or as differentiated therefrom. Thus it is only those self-referential assertions, the truth-claims of which include in some way the truth-conditions of their object-sentences, that is, self-referential assertions concerning truth-conditions (either directly or through assertions of truth or falsity), which will be identified in their truth-conditions with their object-sentences and thereby rest indefinite in those very truth-conditions, due to the unending contingency of the latter on infinitely generated object-sentences.

Let us not consider assertions of the form "p' yields a falsehood when appended to a quotation of itself", where p is a predicate lacking an argument (e.g. 'is a string of German phonemes'). The truth or falsity of propositions such as this is contingent upon the definiteness of the truth-conditions of the sentence "p' p" (e.g. "Is a string of German phonemes' is a string of German phonemes") generated from the embedded predicate, the non-obtaining of which in the world forms the truth-conditions for that original assertion, unless those sentences are identified. Let us take an assertion of this form which Quine considers a genuine antinomy, "Yields a falsehood when appended to a quotation of itself' yields a falsehood when appended to a quotation of itself." The "p' p"-form assertion generated by this assertion may be identified with the generating assertion or not. If it is identified, then the truth-conditions are identical and the definiteness of the generating assertion is contingent upon the definiteness of the generated sentence. However, the truth-conditions of the first generated assertion involve another generated "p' p"-form sentence with which it must be identical (if it is to be taken as truly identical to the first generating sentence) and thus upon the truth-conditions of which its own truth-conditions are contingent. This contingency of generating sentence truth-conditions on generated sentence truth-conditions is infinitely extended through endless generation of "p' p"-form sentences. Thus, again, no actual paradox is generated as the assertion infinitely regenerates an indeterminate element and thereby fails to fully define truth-conditions.

Russell's paradox, as a problem of natural semantics, allows a similar, if somewhat more complex, solution. Let us take the paradox under the form 'The

description of all descriptions which are not included in their own extensions is included in its own extension.' Obviously we have here a question of inclusion or non-inclusion of this very description in a certain set, and for any description to be included in this set it must not apply to itself, which is to say, it must not be included in the set or extension which it itself defines. Thus we have two criteria which any object must fulfill to be included in the set in question: 1) the object must be a description and 2) it must not apply to itself. Thus, clearly, 'large-nosed bipeds' would be included and 'descriptions in English' would not. It is equally clear that descriptions such as 'descriptions of a particular sort', when spoken by our psychotic, who has no particular sort in mind, cannot be said to be included or not included; the assertion "'Descriptions of a particular sort' is a description of that sort" is not fully definite. Now amongst the set of all descriptions we find 'descriptions which are not included in their own extensions', which, as a description, fulfills the first condition; however, its own extension, which must be definite as to its own inclusion or non-inclusion if it is to be taken as included or non-included in the original set, is defined by precisely those criteria which define the original set, with which set it is, indeed, identical. Thus the definition of the first set with regard to its own description is contingent upon the definition of this second set with regard to its own description, but it should be clear that this second definition is contingent upon yet a third, and so on infinitely, with all sets and all definitions identical and indefinite, leaving 'The description of all descriptions which are not included in their own extensions' indefinite at precisely the point at which it would define the conditions for the truth or falsity of 'The description of all descriptions which are not included in their own extensions is included in its own extension.'

It may be objected here, following Kripke, that it makes no difference if we show single-sentence self-references to be non-paradoxical, as paradoxes might be generated by sentence pairs. Let us take the following pair¹⁰: A) 'B' is true. B) 'A' is false. It may seem that sentence A, for example, has fully definite truth-conditions as it is clearly not identified with its object-sentence (i.e., clearly, "'B' is true" ≠ "'A' is false"); however, it is identified with the object-sentence of its object-sentence and, just as there is an element of indeterminacy in the original metalinguistic sentence, there is in this second object-sentence such an element of indeterminacy, which must ultimately be determinate if the truth-conditions of the metalinguistic sentence are to be fully determinate, for the determinacy of the truth-conditions of the first sentence is contingent upon the determinacy of

the truth-conditions of this third sentence. However, the determinacy of the truth-conditions of this third sentence is itself contingent upon the determinacy of the truth-conditions of the object-sentence of its object-sentence, that is, of some fifth sentence, itself, in turn, contingent, and so on infinitely. Clearly, the same holds for sentence B as well and, hence, neither is fully definite in its truth-conditions.

Related to this is the case of universally quantified assertions about assertions, which gives rise to another 'paradox', the "paradox of grounding in semantics."¹¹ Obviously enough, an assertion of (natural) semantics of the form 'All sentences are x' necessarily generates the observation sentence "'All sentences are x' is x." It does not follow from this, however, that universally quantified assertions of semantics are necessarily undecidable or paradoxical; indeed, it does not even follow that such assertions lack truth-value. First of all, if there is some other sentence 'p' such that "'p' is x" is false, then 'All sentences are x' is itself false. For example, the sentence 'All sentences are false' is not paradoxical, or lacking truth-value; it is simply false.¹² Furthermore, 'is x' may not concern the truth-conditions of embedded sentences whatsoever. For example, 'All sentences are composed of universal semantic units' will generate the observation sentence "'All sentences are composed of universal semantic units' is composed of universal semantic units", but the truth-conditions of the embedded sentence are irrelevant to the truth or falsity of this observation sentence. For any universally quantified assertion about assertions to even seem paradoxical, its predicate must concern the truth-conditions of all embedded sentences of all observation sentences and it itself must be definitely unfalsified by all assertions excepting itself. However, only predicates of truth-value fulfill the first condition and 'All sentences are true', 'All sentences are false' (as we have noted) and even 'All sentences are either true or false' or 'All sentences fully define truth-conditions' are false, and therefore fail to fulfill the second condition (while assertions such as 'All sentences the truth-conditions of which obtain in the world are true' or 'All and only sentences with determinate truth-conditions are either true or false' are definitional and thereby have no observation sentences and no truth-conditions, definite or not).

Thus natural semantics is by no means a necessarily paradoxical undertaking; indeed there are not even any universally quantified assertions about assertions which are non-true/non-false due to a failure to fully determine truth-conditions. Furthermore, other reputed paradoxes deriving from direct sentential self-reference or from mutual reference of sentence pairs are

neither true nor false (and, paradoxical) as they fail to fulfill conditions. Thus we see that by careful truth-conditions paradoxes of self-reference are eliminated without violating any of our theoretical criteria."

NOTES

¹The Master Thinkers, trans. Brian Pearce (New York: Harper and Row, 1980), p. 13.

²Conjectures and Refutations: The Growth of Scientific Knowledge (New York: Harper and Row, 1963), p. 309. A similar, 'prosentential' account, such as that of Dorothy Grover, is equally problematic, as John Pollock has pointed out ("Inheritors and Paradox," and "The Liar Strikes Back," respectively, Journal of Philosophy, LXXIV, No. 10).

³"A New Approach to the 'Liar'," Danish Yearbook of Philosophy, vol. 5, pp. 8-9.

⁴In The Paradox of the Liar, ed. Robert L. Martin (New Haven: Yale University Press, 1970), p. 75.

⁵Objective Knowledge: An Evolutionary Approach (Oxford: Clarendon Press, 1972), p. 304.

⁶It should be clear that I hold to a Tarski-type notion of truth; specifically, I assume that there are certain intended articulate descriptions (infinite in number and concerning infinitely many sets of intentional objects) each of which may be said to apply in all particulars to the world, not describing the world completely, but giving an accurate partial description, and certain other intended articulate descriptions (also infinite in number) which may be said, in one or another particular, not to apply to the world. (In the following, I endeavor to demonstrate that paradoxical and undecidable self-referential propositions and sets of propositions belong to neither group.)

⁷Or, more exactly, conditions defined by the intention of the asserting subject. On meaning as intention, see my "Meaning, Intention, and Mind," The Modern Schoolman, LIX (May, 1982).

⁸This has been noted in part by Gilbert Ryle in his brief but highly suggestive discussion of the liar's paradox; see "Heterologicality" in Philosophy and

Analysis, ed. Margaret MacDonald (Oxford: Basil Blackwell, 1954), p. 52.

'The Ways of Paradox (New York: Random House, 1966).

¹⁰I choose these rather than Kripke's more obviously 'realistic' examples as Kripke's examples suffer problems of temporal range and therefore import irrelevant problems about the truth-value of propositions referring to the future. For Kripke's original examples, see "Outline of a Theory of Truth," Journal of Philosophy, LXXII, pp. 669ff.

¹¹See Hans Herzberger, "Paradoxes of Grounding in Semantics," Journal of Philosophy, LXVII, No. 6. My discussion applies only indirectly to Herzberger's arguments as I am concerned with natural, and not formal, semantics; however, as I have noted above, I believe natural semantics to the proper domain of any discussion of the actual truth or falsity of propositions.

¹²This has been noted, in a different context, by A. N. Prior. See his "On a Family of Paradoxes," Notre Dame Journal of Formal Logic, II, No. 1, p. 27.

¹³I am indebted to Professor Donald Davidson for penetrating comments on an earlier, and quite different, version of this paper. Portions of this essay were included in a series of lectures, entitled "Meaning and Possible Worlds," delivered by the author at the University of Ljubljana, in June 1982.