Semantical Paradox*

Tyler Burge

Frege remarked that the goal of all sciences is truth, but that it falls to logic to discern the laws of truth. Perceiving that the task of determining these laws went beyond Frege's conception of it, Tarski enlarged the jurisdiction of logic, establishing semantics as truth's lawyer.[1]

At the core of Tarski's theory of truth and validity was a diagnosis of the Liar paradox according to which natural language was hopelessly infected with contradiction. Tarski construed himself as treating the disease by replacing ordinary discourse with a sanitized, artificial construction. But those interested in natural language have been dissatisfied with this medication. The best ground for dissatisfaction is that the notion of a natural language's harboring contradictions is based on an illegitimate assimilation of natural language to a semantical system. According to that assimilation, part of the nature of a "language" is a set of postulates that purport to be true by virtue of their meaning or are at least partially constitutive of that "language". Tarski thought that he had identified just such postulates in natural language as spawning inconsistency. But postulates are contained in theories that are promoted by people. Natural languages *per se* do not postulate or

[1] Gottlob Frege, "The Thought", in E. D. Klemke, ed., *Essays on Frege* (Urbana: Univ. of Illinois Press, 1968). Alfred Tarski, "The Concept of Truth in Formalized Languages", in *Logic, Semantics, Metamathematics*, J. H. Woodger, trans. (New York: Oxford, 1956). Nothing that I say will depend heavily on precisely how one views the scope of logic.

assert anything.[2] What engenders paradox is a certain naive theory or conception of the natural concept of truth. It is the business of those interested in natural language to improve on it.

Another just ground for dissatisfaction with Tarski's diagnosis is that it does not deal with various intuitions associated with the natural notion of truth. A philosophically satisfying theory must administer to these intuitions. Post-Tarskian treatments of the paradoxes, with a very few exceptions, have shared this second failing in greater or lesser degree. Although the motivation for these treatments is purportedly to provide a more natural or intuitive account of truth, they tend at least implicitly to place a higher premium on technical ingenuity than on intuitive adequacy. There results a variety of emendations of classical semantics without any thoroughly developed motivation. Important intuitive aspects of the paradoxes are usually left untouched.

My objective is an account of the "laws of truth" whose application accords as far as possible with natural "pre-theoretic" semantical intuition. Under these laws I shall be prosecuting the Grelling and Liar Paradoxes. The puzzles of Berry and Richard present slightly different cases; somewhat further afield are the epistemic and modal paradoxes.[3] But, although each of these cases deserves a special hearing, the basic outlook of this paper is intended to carry over to them.

I.

Tarski's analysis of the Liar allowed three escape routes. One could deprive the language of the means to name its own sentences. One

[2] This point runs far deeper than these brief remarks indicate. The problem of diagnosing Tarski's mistake has been most extensively and carefully discussed by Hans Herzberger, in "The Logical Consistency of Language", in J. A. Emig, J. T. Fleming, and H. M. Popps, eds., *Language and Learning* (New York: Harcourt, Brace, & World, 1966); and "The Truth-conditional Consistency of Natural Languages", *The Journal of Philosophy*, LXIV, 2 (Feb. 2, 1967): 29–35. I do not find Herzberger's diagnoses, which attribute to Tarski a rather simple inconsistency, entirely convincing either intuitively or textually. In my view, Tarski's error is more like a category mistake than an inconsistency.

[3] Cf. my "Buridan and Epistemic Paradox", *Philosophical Studies*, XXXIV, 1 (July 1978): 21–35; David Kaplan and Richard Montague, "A Paradox Regained", *Notre Dame Journal of Formal Logic*, I, 3 (July 1960): 79–90; and Richard Montague, "Syntactical Treatments of Modality with Corollaries on Reflexion Principles and Finite Axiomatizability", *Acta Philosophica Fennica*, XVI (1963): 153–67.

could limit the practice of asserting the results of substituting, for each sentence of the language in which the antinomy is constructed, a name of the sentence for 'X' and the sentence itself for 'P' in the truth schema

$$X \text{ is true if and only if } P.$$

Or one could restrict classical rules of transformation or alter semantical assumptions underlying classical logic. The first route is not viable for anyone who wishes to account for natural language or deductive reasoning in mathematics. Tarski chose the second and rejected the third out of hand.

Roughly speaking, Tarski's approach involves attaching numerical subscripts to 'true' (yielding a hierarchy of predicate constants) and treating as ill-formed any attempt to predicate $\ulcorner true_m \urcorner$ of a sentence containing $\ulcorner true_n \urcorner$, $n \geq m$. Thus a truth$_m$ schema is restricted to apply only to sentences containing no predicate $\ulcorner true_n \urcorner$, $n \geq m$. This requirement effectively blocks the derivation of contradiction, and it has become standard among mathematical logicians. But philosophers writing on the subject, with a few exceptions, have tended to agree with Tarski in denying the applicability of the approach to natural language.

Criticisms of Tarski's construction as a resolution of the natural-language paradoxes have taken several forms. It has been held that Tarski gave little motivation for the hierarchy except as an obstruction to contradiction and provided little insight into the use of the term 'true'; that 'true' is univocal, whereas Tarski fragments the notion of truth into infinitely many predicate constants; that there are global applications of 'true'—"every proposition is either true or not"—that Tarski's theory cannot represent; that paradoxical sentences are not ungrammatical and sometimes lead to difficulty not because of anything odd about their meaning, but because of empirical facts; and finally that there are cases of perfectly normal semantical evaluation which are pronounced abnormal from Tarski's viewpoint. All these criticisms have some merit as applied to Tarski's own theory, though I think that none reach quite so deeply as their authors have supposed.

The past fifteen years or so have seen a swell of support for combining the second avenue of escape with the third. That is, restrictions on the truth schema are conjoined with alterations of classical transformations or classical semantical assumptions. Such

approaches always include rejection of the principle of bivalence. The intuitive motivations for this strategy have been various, and articulated only sketchily if at all. For reasons of space, my discussion of them will be correspondingly sketchy.

One motivation is a desire to integrate a solution to the semantical paradoxes with a theory of presupposition associated with Frege and Strawson.[4] It is certainly desirable to mark the difference between presupposition failure and falsity, ordinarily so-called, in our theory of language. But it is doubtful that adequately marking such a difference requires alterations in the semantics of elementary logic.[5] More important, as will be seen below, the appeal to the alterations occasioned by the Frege–Strawson theory is insufficient to explain certain features of the paradoxes.

A second motivation for altering classical semantics is a desire to assimilate a solution of the paradoxes to a theory of category mistakes.[6] Epimenides' error matches category mistakes in blatancy, but seems to have little else to do with them. Paradoxical statements can be constructed in which the reference of the singular term seems to be the right *sort* of thing for the semantical predicate to apply to— for almost any independently motivated view as to what the right sort is. The relevance of category considerations is thus obscure.

---

[4] Bas C. van Fraassen, "Presupposition, Implication, and Self-reference", *The Journal of Philosophy*, LXV, 5 (Mar. 7, 1968): 136–52; Frege, "On Sense and Reference" in P. T. Geach and Max Black, eds., *The Philosophical Writings of Gottlob Frege* (Oxford: Blackwell, 1966); P. F. Strawson, "On Referring", *Mind*, LIX, 235 (July 1950): 320–44; and *Introduction to Logical Theory* (London: Methuen, 1952).

[5] Cf. H. P. Grice, "Logic and Conversation" in D. Davidson and G. Harman, eds., *The Logic of Grammar* (Encino, Calif.: Dickenson, 1975); Robert Stalnaker, "Presuppositions", *Journal of Philosophical Logic*, II, 4 (October 1973): 447–57, and "Pragmatic Presuppositions", in Milton Munitz and Peter Unger, eds., *Semantics and Philosophy* (New York: NYU Press, 1974); Enrique Delacruz, *Presupposition: Towards an Analysis* (dissertation, UCLA, 1974); Jay Atlas, "Frege's Polymorphous Concept of Presupposition and its Role in a Theory of Meaning", *Semantikos*, I (1975): 29–44.

[6] This approach is one of several implicit in Gilbert Ryle, "Heterologicality", *Analysis*, XI, 3 (January 1951): 61–9. It has been more subtly developed by Robert L. Martin, "Toward a Solution to the Liar Paradox", *Philosophical Review*, LXXVI, 3 (July 1967): 279–311; and "A Category Solution to the Liar" in Martin, ed., *The Paradox of the Liar* (New Haven, Conn.: Yale, 1970). Martin now rests little weight on the category idea. He sees it as subsumable under considerations of presupposition. I would apply remarks similar to those which follow to occasional suggestions that vagueness is the root difficulty. Cf. also the beginning of sec. III.

Moreover, as before, accounting for category mistakes does not clearly require alterations in classical semantics, and such alterations as have been proposed are intuitively inadequate.

A third motivation is a yearning to produce a language in which there is a single truth predicate with constant extension which applies to everything that can be said (truly) in the language.[7] I shall argue that this ideal overlooks certain simple intuitions about truth. It is also doubtful that the ideal is technically attainable without exorbitant intuitive costs. This latter point, however, reaches beyond our present treatment.

Consistent, nonbivalent logics with a univocal truth predicate are certainly constructible. But no such logic, insofar as it assumes a truth predicate with a constant extension, has given a plausible account of the semantical paradoxes. This is because of a family of problems that have become known as the "Strengthened Liar". *The Strengthened Liar* (perhaps better called "The Persistent Liar") is really the original Liar reiterated for the sake of those who seek to undercut paradox primarily by appeal to a distinction between falsehood and some other kind of truth failure. Failure to resolve the Strengthened Liar is not a difficulty of detail or a mere drawback in a solution. It is a failure to account for the basic phenomenon. Any approach that suppresses the liar-like reasoning in one guise or terminology only to have it emerge in another must be seen as not casting its net wide enough to capture the protean phenomenon of semantical paradox.

The Strengthened Liar in its simplest form is this. If we analyze

(β)   (β) is not true

as being neither true nor false, then it intuitively follows that the sentence displayed is not true. But the sentence displayed is (β). So it seems to follow that (β) is not true after all. We have now apparently asserted what we earlier claimed was neither true nor false. Moreover, the assertion that (β) is not true would seem to commit us to asserting that '(β) is not true' is true, contrary to our original

---

[7] Cf. van Fraassen, op. cit.; Martin, op. cit.; Brian Skyrms, "Return of the Liar: Three-Valued Logic and the Concept of Truth", *American Philosophical Quarterly*, VII, 2 (April 1970): 153–61; and Hans Herzberger, "Truth and Modality in Semantically Closed Languages" in *The Paradox of the Liar*, op. cit.

analysis.[8] It is important to see that this informal reasoning is entirely intuitive.

Although the problem is well known, truth-value-gap theorists—theorists who propose to handle the paradoxes primarily by denying bivalence and who espouse a semantics with "truth-value gaps"—have had little illuminating to say about it. For example, in a technically elegant and critically acute paper, Saul Kripke raises the issue almost as an afterthought.[9] He admits that the Liar sentences are not true in his gap-containing object language and that this point cannot be expressed in that language. And he goes on to suggest a further truth predicate in a bivalent metalanguage. But, since Kripke's admission is couched in natural language, the proposal in terms of truth-value gaps *a fortiori* does not cover (at least one use of) 'true' in natural language. In short, an account of truth in the metalanguage—and in natural language—is still needed. Since the

account envisioned is apparently something like Tarski's, the explanatory value of the gaps is unclear.

There have been three types of response among those truth-value gap theorists who have discussed the Strengthened Liar in detail. One is simply to reject the remark that the sentence displayed (under any description) is not true. The rationale goes: one must take seriously one's claim that 'true' is undefined or "gappy" for some arguments; the Strengthened Liar simply seduces one into forgetting this insight.[10] Sometimes the point is made in terms of the pathological statement's being "indeterminate" in some non-epistemic sense, terms that seem unusually promising of intuitive obscurity.

Implicit in this rationale is the view that the metalanguage for the gap-containing object language must itself have the same gaps the object language has. The response manifests a courageous and admirable methodological consistency, but is hardly credible. Informally and intuitively, to say that a sentence is neither true nor false is to imply that the sentence is not true.

Claiming that in the problem sentence the truth predicate is undefined or its application indeterminate does not help matters. For one may still reason that, if a sentence's predication is undefined or indeterminate, then the sentence is not true. This reasoning may or may not involve a broadening of the domain of discourse or a sharpening of the extension of 'true'. But it is informally quite intuitive. And the response does nothing to account for it.

Even apart from these problems, the response merely encourages the paradox to assume a different terminology. This can be seen by considering versions of the Strengthened Liar adapted to fit the very words of the response:

(i) (i) is either false or undefined
(ii) (ii) is not determinately true.

After claiming that (β) is neither true nor false (or "bad" in some other sense), the gap-theorist must still face a precisely analogous Strengthened Liar tailored to his favorite description of the gaps.

---

[8] The general problem was first articulated by A. P. Ushenko, *The Problems of Logic* (London: Allen & Unwin, 1941), pp.. 78 ff. Cf. also W. W. Rozeboom, "Is Epimenides Still Lying?", *Analysis*, XVIII, 5 (April 1958): 105–13; and van Fraassen, op. cit. The term 'Strengthened Liar' is van Fraassen's. He may apply it somewhat more narrowly than I do.

It has been suggested that there is another paradox for truth-value-gap theories, which centers on falsity. Take '(a) is false'. Suppose (a) is neither true nor false. Then (a) is not false. We have now asserted '(a) is not false'. So we are committed to its truth. '(a) is not false' is the negation of (a). So the negation of (a) is true. But if the negation of (a) is true, then (a) itself is false. (This latter is a principle that even most truth-value-gap theories have accepted.) In my view, this reasoning is indeed a problem for truth-value-gap theories. But it is slightly more complicated than the reasoning I fix upon. The theory I develop will, however, handle this version of the "Strengthened" Liar in an obvious way, regardless of whether one counts 'false' as equivalent to 'is well-formed and is not true'.

[9] "Outline of a Theory of Truth", reprinted in this volume, pp. 79–81. Kripke remarks that certain technical terms (like 'paradoxical', 'grounded') do not occur in natural language "in its pristine purity". I see no interesting or clear distinction between terms reflectively introduced into natural language for unchallenged explanatory purposes and terms that slip in by other means. But, even if there were a relevant distinction, the theoretical situation would not be altered, since 'true' causes the difficulty unaided by more technical terms. Kripke further suggests that 'true' as it appears in his metalanguage corresponds to a 'true' in natural language different from (more technical than?) the 'true' that the predicate in his gap-containing object language corresponds to. The idea is that the former use of 'true' arises at a "later stage in the development of natural language" as a result of reflection on the use of 'true' in the gap-containing language. I find this unsatisfying, since I see no intuitive or philological basis for the claim that the natural-language 'true' changes its sense or logic as a result of reflection. Kripke's suggestion is similar to the view in van Fraassen, "Truth and Paradoxical Consequences", in *The Paradox of the Liar*, op. cit.

[10] J. L. Mackie, *Truth, Probability and Paradox* (New York: Oxford, 1973), pp. 290–5; Martin, "A Category Solution", op. cit., pp. 92, 96. The indeterminacy idea that follows is Mackie's. On this issue it is a mistake to get too wound up in purely formal issues regarding the difference between choice negation and exclusion negation. There is clearly a use of negation in English which acts enough like exclusion negation to cause the problem, and we may start the paradox with this use.

A second response is to place restrictions on substitutivity of identity to block the move from remarking that the displayed sentence is not true to asserting that (β) is not true.[11] The response is palpably *ad hoc*. And it comes to feel more so when one considers that paradoxes can be constructed without using substitutivity of identity, or even singular terms. Brian Skyrms has proposed to meet some of these problems by restricting universal instantiation and by denying the validity of relettering bound variables. There is no evident unifying conception behind these restrictions—no basis other than minimum mutilation on which to choose restrictions as difficulties arise. Even as it stands the logic is cumbersome and unintuitive.

In addition, there are intuitions that the approach does not capture. I adapt an example given by Buridan:

(A) Suppose Plato at time t_1 says, "What Aristotle says at t_1 is not true" and Aristotle at time t_1 says "What Plato says at t_1 is not true". Since the two sentences are related to one another in exactly the same way, there can be no reason for saying that one is true and the other is not. But if we assume the naive truth schema, if one sentence is true the other must not be. It follows that neither statement is true—both are pathological and will not fit the naive schema. But intuitively, a third party who goes through the appropriate reasoning could well say, "What Plato says at t_1 is not true" (as well as "What Aristotle says at t_1 is not true") without saying anything paradoxical.

The third party uses the same singular term ('what Plato says at t_1') that Aristotle did (and the term could obviously be made non-indexical). Paradoxicality does not seem to depend purely on what mode of reference is used.[12]

There is another argument that the mode of reference in the Liar sentence is not the real source of paradox:

(B) Suppose I conduct you into a room in which the open sentence type 'it is not true of itself' is written on a blackboard. Pointing at the expression, I present the following reasoning: Let us consider it as an argument for its own variable or pronoun. Suppose it is true of itself. Then since it is the negation of the self-predication of the notion of *being true of*, it is not true of itself. Now suppose it is not true of itself. Then

[11] Skyrms, "Return of the Liar", op. cit.; and "Notes on Quantification and Self-reference", in *The Paradox of the Liar*, op. cit.

[12] Jean Buridan, *Sophisms on Meaning and Truth*, Kermit Scott, trans. (New York: Meredith, 1966), pp. 200 ff. A similar point can be made if A says 'what P says is not true' and P says 'what A says is true'.

since it is the negation of the self-predication of the notion of *being true of*, it is true of itself. In response you suggest that it is undefined for itself, from which we conclude that it is not true of itself. But then, I ask, why have we not made the same predications we were just criticizing? If we have, we seem committed to its being true of itself after all.

Here we have a variant of the Grelling paradox, analogous to the Strengthened Liar. Yet we have used only one mode of reference—the pronoun 'it' whose antecedent is a demonstration. I conclude that this second response to the Strengthened Liar is misguided.

Examples (A) and (B) leave rather little room in which to maneuver. In the passage from recognizing the relevant expression as pathological to noting *in the very words of the expression* that it is not true (or not true of itself) and then counting *this* remark true (or true of itself), we have an intuitive change of evaluation. The change does not seem attributable to the relevant singular expressions. That appears to leave a shift in the extension of the truth predicate and a shift in the applications of negation as possible sources of the change in evaluation.[13]

A third response by gap theorists fixes on negation.[14] The idea is that when we say of 'it is not true of itself' that it is not true of itself, we are using a broader sense of negation than the sense used in the pathological occurrence of 'it is not true of itself'. This idea seems the most plausible of those discussed here. But it does not promise a unified account of the paradoxes. There are semantical sentences lacking negation ('This is true') which, though not paradoxical, are pathological in a way intuitively analogous to the Liar. Moreover, it

[13] Actually there are other possibilities. Two interesting ones are suggested by Charles Parsons, "The Liar Paradox", reprinted in this volume. Parsons develops the ideas that there may be shifts in the domain of discourse or shifts in the extension of the notion of *expressing*—a relation between sentences and propositions. A thorough discussion of these ideas would take us too far afield. Suffice it here to say that in (A) and (B), we seem to be speaking of an English expression, standardly construed, throughout the example. So the domain of discourse appears to remain constant. As for shifts of expression, one needs an argument for the universal applicability of the model which relates sentences to nonlinguistic intensional entities. It seems natural to think that paradox may arise even where the truth bearers are sentences (as interpreted in a context) or tokens. Moreover, many of the points we make below will apply by analogy to an "expression" relation. On the latter of Parson's approaches we do not yet have a systematic account of how truth value is determined by the semantical or pragmatic roles of sentential components.

[14] Frederic B. Fitch, "Universal Metalanguages for Philosophy", *Review of Metaphysics*, XVII, 3, 67 (March 1964): 396–402. Fitch does not actually discuss the present problem. But his treatment of a related problem involves this approach.

is known that semantical paradoxes can be produced *without* negation, using only the truth schema, *modus ponens*, and the inference rule: from $A \supset . A \supset B$ to infer $A \supset B$.[15] These negation-less paradoxes can be cast into strengthened form by informal reasoning about the material conditional. Now one might use the restrictions on the truth schema, which all gap theorists appeal to, to treat the "ordinary" paradoxes (and pathologies like 'This is true'), and a hierarchy of negations (and material conditionals!) to deal with the strengthened versions. But such an approach, though technically feasible, promises little philosophical illumination. The semantical paradoxes are remarkable in their similarity. The Strengthened Liar does not appear to have sources fundamentally different from those of the ordinary Liar. What is wrong with the proposed account is that it gives no insight into the general phenomenon of semantical pathology and offers instead a hodge-podge of makeshift and merely technical remedies. A theory of semantical paradox should focus on semantical notions.

The Strengthened Liar indicates that whatever other virtues truth-value gaps may have, they do not themselves mitigate the force of paradox. Indeed, they do little more than mark, in a specially dramatic way, the distinction between pathological sentences and sentences that are ordinarily labeled "false".

We have always had reason to distinguish '*x* is false' from '*x* is not true'. Nonsentences (or even open sentences apart from a context of application) are obviously or categorically not true. But one has no inclination to call them "false". Tarski regarded 'is false' as amounting to 'is a closed sentence that is not true'. Truth-value-gap theorists have seen this identification as unnatural, since some (closed) sentences seem to go wrong in ways that are deeper or prior to falsity. I am sympathetic with this viewpoint, at least as directed toward natural-language 'false'. If for the moment we ignore (with Tarski) qualifications needed for indexical sentences, we may see 'false' as appropriately applied to a proper subset of closed sentences that are not true—a subset meeting certain further *pragmatic* conditions. Such a view is compatible with retention of a highly general interpretation of the business of semantics, as giving laws for sorting true sentences, interpreted in a context, from those which are

---

[15] Haskell B. Curry, "The Inconsistency of Certain Formal Logics", *Journal of Symbolic Logic*, VII, 3 (September 1942): 115–17; P. T. Geach, "On *Insolubilia*", *Analysis*, XV, 3 (January 1955): 71/2.

not. Tarski himself rarely used the term 'false', sticking mostly with 'true' and negation. From this viewpoint, truth is not strictly the *value* of a function, as Frege held. It is rather seen on the model of a property that the relevant truth bearers, and everything else, either have or lack. This view is efficient as well as traditional. To treat the paradoxes, there is no need to give it up. Restricting the truth schema is the essential curative.

We have seen that some gap theorists envision combining truth-value gaps with some sort of hierarchy. And I believe that a move in this direction is necessary even to begin to cope with the intuitive evidence. But such a move seems to draw in its wake all or almost all the criticisms, mentioned at the beginning of this section, that gap theorists have leveled at Tarski's hierarchy. In view of the preceding, it would seem simpler to worry less about the gaps and rethink the hierarchy.

II.

In all the variants of the Strengthened Liar so far discussed, we started with (*a*) an occurrence of the Liar-like sentence. We then reasoned that the sentence is pathological and expressed our conclusion (*b*) that it is not true, in the very words of the pathological sentence. Finally we noted that doing this seemed to commit us to saying (*c*) that the sentence is true after all. Example (A) is especially striking. We seem to pass without intuitive difficulty from the hopeless tangle that Plato and Aristotle have got them-selves into to the reasoned comment of the third party. Yet the third party uses the same sentence that was used by one or both of the tangled. The first task of an account of semantical paradox is to explicate the moves from (*a*) to (*b*) and from (*b*) to (*c*).

Most recent accounts have either ignored such reasoning as the above or sought simply to block it by formal means.[16] I think a more satisfying approach is to interpret the reasoning so as to *justify* it. The intuitiveness of the informal reasoning that generally occurs in the throes of paradox has been obscured by a concentration on simple, obviously perverse examples. I think it well to review a case

---

[16] There are exceptions. See A. N. Prior, "On a Family of Paradoxes", *Notre Dame Journal of Formal Logic*, II, 1 (January 1961): 16–32. Cf. also Herzberger, "Truth and Modality in Semantically Closed Languages", in *The Paradox of the Liar*, op. cit., esp. p. 31; and Parsons, op. cit.

(derived from Prior; cf. note 16) that is formally similar but less bizarre than the preceding.

(C) Suppose a student, thinking that he is in room 10 and that the teacher in room 9 is a fraud, writes on the board at noon 8/13/76: (*a*) 'There is no sentence written on the board in room 9 at noon 8/13/76 which is true as standardly construed'. Unfortunately, it being Friday the 13th, the student himself is in room 9, and the sentence he writes is the only one on the board there-then. The usual reasoning shows that it cannot have truth conditions. From this, we conclude that it is not true. But this leads to the observation that (*b*) there *is* no sentence written on the board in room 9 at noon 8/13/76 which is true as standardly construed. But then we have just asserted the sentence in question. So we reason (*c*) that it is true.

Before interpreting the reasoning in detail, I shall interpret it in summary. In the moves from (*a*) to (*b*) to (*c*) in example (C), there seems to be no change in the grammar or linguistic meaning of the expressions involved. This suggests that the shifts in evaluation should be explained in pragmatic terms. Since there is a shift from saying that the relevant sentence is not true to saying that the same sentence *is* true [(*b*) to (*c*)]—a shift in truth value without change of meaning—there is an indexical element at work.

The indexicality is most plausibly attributed to the truth predicate. As we have seen, there may or may not be a singular term in the examples, and any such singular term may or may not be indexical. Negation is not a regular feature in semantical pathology. Thus indexicality in the semantical predicates seems to be the natural alternative. The central idea in accounting for the move from (*b*) to (*c*) will be to interpret 'true' as contextually shifting its extension. In (*b*) we claim that the original paradoxical sentence (at the relevant occurrence) is not true—given the context of application of the occurrence of 'true' within it. Let us mark this occurrence as 'true$_i$'. But from a broader, or subsequent application of 'true', undertaken in (*c*), the sentence (at the relevant occurrence) is true (true$_k$)—since, in effect, it says it is not true$_i$; and it is not.

This explanation, which is sketchy and somewhat misleading, will be developed in sections IV and V. It is worth noting now, however, that the original sentence as interpreted at the relevant occurrence is not granted truth$_i$ conditions. That is, we should not insert the sentence so interpreted into the truth schema for 'true' construed as it occurs in that sentence ('true$_i$') and assert the resulting bicon-

ditional. All the solutions agree on some such restriction. The truth schema for 'true$_k$', however, may take the original sentence [interpreted as it is in step (*b*) and containing 'true$_i$'] as instance. The sentence is *not true$_i$*—not because its truth$_i$ conditions are not fulfilled, but because it has no truth$_i$ conditions. But it does have truth$_k$ conditions and indeed is true$_k$.

What of the move from (*a*) to (*b*)? For reasons that I will give in section III, I think that it does not strictly involve an indexical element in the sentences themselves. Rather, there is a shift in certain implicatures pragmatically associated with the sentence occurrences.

The relevant implicature is that *sentences being referred to or quantified over are to be evaluated with the truth schema for the occurrence of 'true' in the evaluating sentence*. In the paradoxical occurrence (*a*) the sentence referred to or quantified over is the evaluating sentence itself [as interpreted in occurrence (*a*)]. So it is implicated that that sentence is to be evaluated with a truth$_i$ schema. The sentence is shown to be pathological by taking the implicature seriously and applying the relevant schema. The implicature is scrutinized in the reasoning that shows that the application of the truth$_i$ schema leads to absurdity. When the same sentence is reasserted in (*b*) (and this sentence occurrence also lacks truth$_i$ conditions), the implicature has been canceled. The sentence [as it occurs in both (*a*) and (*b*)] may be evaluated in a broader semantical context (that of 'true$_k$'). But the reasoning behind the assertion of the sentence in step (*b*) is precisely that the truth$_i$ schema does not apply.

A simplified summary of the interpretation of examples like (C) follows:

step (*a*): (I):   (I) is not true
                     Represented as: (1): (1) is not true$_i$
                     Implicature: (1) is evaluated with truth$_i$ schema.
step (*b*):   (I) is not true (because pathological)
                     Represented as: (1) is not true$_i$
                     The implicature of step (*a*) is canceled.
step (*c*):   (I) is true after all
                     Represented as: (1) is true$_k$
                     Implicature: (1) is evaluated with truth$_k$ schema.

I have taken (1) to be a sentence interpreted in a context. But one may just as well take it to be a token.

III.

In this section I want to discuss the reasons for interpreting the move from (a) to (b) in terms of implicature. A natural reaction to example (C) is to say that the token of the sentence that is in room 9 is pathological or fails to express a proposition, whereas tokens outside room 9 are true.[17] This reaction may be interpreted as compatible with our viewpoint. The failure to express a proposition might be taken to *consist in* failure to be assigned truth conditions by the truth schema that is implicated to be appropriate.

This way of seeing the matter provides a partial rapprochement between our view and truth-value-gap approaches. We agree that in a limited sense not every well-formed sentence interpreted as used in a context is true or false, or even has truth conditions: Certain such sentences fail to be true (given an indexical use of 'true'—call it 'true$_i$') not because their truth$_i$ conditions are not fulfilled, but because they have none. Where our view differs from the truth-value-gap approaches is in its observation that this "failure to have truth conditions" (truth$_i$ conditions) is not an absolute affair. The same sentence interpreted in a context (or, if you prefer, same token) that lacks truth$_i$ conditions may, indeed will, have truth$_k$ conditions, and can be evaluated as true$_k$ or false$_k$. The plausible intuition that a given sentence, interpreted in a context, (or sentence token in room 9) has gone "bad" in a sense prior to falsity depends on the fact that the truth$_i$ schema that is pragmatically implicated to be appropriate for evaluating the sentence is not applicable to it. (At this point, the reader need not worry about the precise significance of the subscripts on 'true'. That will be explained in sections IV and V. It is enough to see them as marking contextually different applications of 'true' which yield different extensions for the indexical predicate.)

There are two important restrictions on any intuitive claim that the sentence (or token) occurring in room 9 in example (C) does not express a proposition or statement. First, the precise sense of 'proposition' must be explicated. For, if the term is taken in certain traditional senses, the claim will be mistaken. One cannot reasonably say that all paradox-producing sentences fail to have a

meaning or sense. Such sentences can be used in informal reasoning; we can express their content via semantic ascent; and their paradoxicality sometimes results from empirical facts, rather than anything intrinsic to their meaning. Further, such sentences can seemingly express reasonable beliefs or thoughts, even on occasions where they lead to paradox. This point is subject to further discussion elsewhere.[18] But, dogmatically speaking, we could have imagined the student in example (C) thinking about thoughts rather than writing about sentences. So 'failure to express a proposition' is not plausibly taken to mean 'failure to express something that could be believed'. Further, we are in the process of arguing that 'failure to express a proposition' need not mean 'failure to express something true or false', in any absolute sense of 'failure' or 'express'. Lacking some explication of 'proposition', then, the claim will be empty. Moreover, as it stands, the claim does not appear to touch the Richard paradox or the paradoxes of grounding.

Second, the claim does not obviate the need to explain the semantic or pragmatic mechanism whereby a given sentence changes from "not expressing a proposition" to being true. What is it about the use or meaning of the sentence (or its components) that accounts for the shift?

A tempting construal is to take the move from (a) to (b) to involve a shift in extension via some sort of indexicality, presumably in the truth predicate. On this view, the problem sentence as it occurs in (a) and (b) would receive different semantical evaluations: "bad" and "true", respectively. Unfortunately, the interpretation immediately leads to problems. Suppose we make explicit the extension of 'true' as it occurs in the pathological sentence token, by marking it with a subscript: 'true$_i$'. Thus, in example (C) step (a), we have (α) 'No sentence written in room 9 . . . is true$_i$'. The comment in step (b) on (α) would then involve a shift of extension yielding (β) 'No sentence written in room 9 . . . is true$_k$'. But one wants to know whether the sentence in the context marked by (α) is *true$_i$* or not. To say we can't ask (or answer) this question is obscurantist and needlessly mysterious. To say something amounting to "neither" is to head back in the direction of truth-value gaps, which we saw merely postpone the question. I take it as obvious that we should not say that the sentence as it occurs in step (a) represented by (α) is true$_i$. This would lead immediately to contradiction.

[17] Ushenko, "An Addendum to the Note on the Liar Paradox", *Mind*, LXVI, 1, 261 (January 1957): 98; Keith Donnellan, "A Note on the Liar Paradox", *The Philosophical Review*, LXVI, 3 (July 1957); 394–7; C. H. Whiteley, "Let Epimenides Lie!", *Analysis*, XIX, 1 (October 1958): 23/4; Jonathan Bennett, Review, *Journal of Symbolic Logic*, XXXII, 1 (March 1967): 108–12.

[18] Cf. my "Buridan and Epistemic Paradox", op. cit.

Thus we should say that the sentence as it occurs in step (a) represented by (α) is not true$_i$: not true, because it lacks truth$_i$ conditions. This is in effect what we do say in step (b). So the occurrence of the problem sentence in step (b) can be represented by (β'), 'No sentence written in room 9 ... is true$_i$'. Step (b) is thus reasonably seen as answering the question of whether (α) in step (a) is true$_i$ or not. Now it is difficult to see how there could be a shift of truth value or semantical evaluation between (α) and (β'), since they are one and the same!

Actually, this remark (though I think it is correct) prejudices the issue slightly since we have not fully formalized the occurrences of the problem sentence in steps (a) and (b). One might think, for example, that, since the occurrence in step (a) is in some broad sense self-referential while the occurrence in (b) is not, there must be some difference in formalization. (Strictly speaking since the problem sentence is and quantifies over a type not a token, it is self-referential in both (a) and (b); but there is an implicated self-referential element, not affecting formalization, which is present in step (a) but not in step (b)—which I shall identify shortly.) One might think that such a difference would reveal a difference of semantical evaluation.

Undermining this thought takes a further argument. Here it is. We have agreed that 'true' does not shift its extension between steps (a) and (b). But the quantifier phrase [which could be simplified to a singular term, as in examples (A) and (B), section 1] does not, or need not, shift its domain or extension either. We can regard ourselves as making reference to a single sentence, interpreted in a contextually determined way, in both step (a) and step (b). (We could also revise the example so that we make reference to a single sentence *token* in room 9. Compare note 13.) So whatever formalizations one gives of the occurrences of the problem sentence in steps (a) and (b) respectively (even if the formalizations are not identical as I think they should be), those formalizations have the same component referents or extensions: they are extensionally isomorphic. So either there is no genuine shift of truth value or semantical evaluation between the formalizations [and between the occurrences of the natural-language problem sentence in (a) and (b)], or 'true' taken with a contextually fixed extension ('true$_i$') is nonextensional—producing truth bearers with different truth values when applied to terms with the same extensions. But there is no independent reason to regard 'true' as nonextensional. It has usually been taken to be

paradigmatically extensional. Since some cases do not strictly involve loss of self-reference *or even change in the manner of reference* [examples (B) and (C)], the relevant restrictions on extensionality would seem as unmotivated as those associated with Skyrms' proposal (section 1). So it is reasonable to conclude that a change in truth value or semantical evaluation is not strictly involved in the move from (a) to (b) in example (C).

The move should be explicated pragmatically—in terms of change in implicatures or background assumptions on the part of those propounding or interpreting the relevant sentences. In step (a) when the problem sentence is first asserted, the implicature is that the sentence quantified over (satisfying the condition on the quantifier) is to be evaluated by a truth schema containing an occurrence of 'true' with the same extension as the occurrence of 'true' ('true$_i$') in the asserted problem sentence. The sentence quantified over turns out to be the asserted problem sentence, and accepting the implicature leads to contradiction. Nothing in the semantics of the problem sentence changes in step (b) when it is reasserted. The difference is that the implicature of step (a) has been canceled. The problem sentence as it occurs in our assertion in step (b) is not true$_i$ [just as it was in step (a)]. But we no longer expect it to have truth$_i$ conditions. To this degree paradox results from false expectations.[19]

This account explains why there *seems* to be a change of truth value. The sentence at the paradoxical occurrence in (a) is pathologically not true (not true$_i$) under the indexical application of 'true' that is implicated to be appropriate. The sentence at its occurrence in (b) is true (true$_k$) under the application that is there implicated to be appropriate. Thus there is a change from our thinking of the sentence at the first occurrence (a) as pathologically not true$_i$ to thinking of the same sentence at its second occurrence (b) as true$_k$. But the sentence at both occurrences is not true$_i$, and true$_k$.

---

[19] Cf. Grice, "Logic and Conversation", op. cit. Cancelability is the primary mark of implicatures. Grice's other general mark is that they are nondetachable in a certain conditional sense. This criterion is vague. But I believe that the relevant implicatures are detachable only insofar as Grice's condition is not satisfied; the manner of expression typically is crucial to the self-reference, hence to the calculation of the implicature.

IV.

The move from (b) to (c) in our examples is a shift from taking a sentence under a contextually determined interpretation as pathologically not true in taking the same sentence under the same interpretation as true. I have proposed to explain such shifts by regarding semantical predicates as indexical. The relevant notion of indexicality must be explicated from two viewpoints, structural and material. We shall take the structural viewpoint first.

The language I espouse as a model of natural language (at least for present purposes) has as its underlying logic standard first-order quantification theory. We place no general restrictions on quantification. Within this language occurs an indexical predicate 'satisfies', which has the usual formal relation to 'true'. (I shall henceforth speak informally only of 'true', and assume for simplicity's sake that no other semantical predicates occur.) These predicates are indexical in the sense that their extensions are not fixed, but vary systematically depending on their context of use. Thus the predicates are not strictly constants, though they may be and often are treated as such for a fixed context. They are not variables either, since we do not quantify over them. Though certain higher-order logics that make special ramification provisions (e.g., ramified type theory) do not quantify into the place of these predicates, these theories do not seem to model natural language perspicuously. Thus 'true' is a schematic predicate. In a given context 'true' takes on a specific extension, and in that context we can represent 'true' with a predicate 'Tr' (or 'Sat') subscripted numerically. The use of *numerical* subscripts is a matter we shall discuss shortly. How a subscript is established in a context is the "material" side of indexicality (section v).

There is considerable agreement that the semantical and set-theoretic paradoxes depend partly on the fact that *truth* and *set* are derivative notions. A sentence like (α) "(α) is true", which is intuitively pathological in the same way that the Liar sentence is, is pathological because (one feels) nothing is stated that can be evaluated as true. Something independent of the evaluation must be established before normal evaluation is possible. Similarly, the notion of a set's containing itself as a member is (to many) pathological because sets are (often) conceived as collections of entities. To be collectible, the members must exist independently of the set. Tarski's language-levels may be regarded as a means of

expressing derivativeness. Part of the intuitive meaning of our subscripts on 'true' is implicit in Tarski's construction. Actually, Tarski's own construction is formally analogous to only one of several constructions—and not the most plausible one—which I shall develop from our indexical viewpoint. But it has the advantage of familiarity; so I shall expound it (or rather its analogue) first.

The basic idea behind all the constructions is to define a notion of a pathological$_i$ sentence. (From here on, I speak of *sentences*, understanding them to be *sentences as interpreted in a context*.) Then we claim that pathological$_i$ sentences are *not true$_i$*, and assert all and only instances of the truth$_i$ schema got from substituting sentences that are not pathological$_i$.[20] A sentence that is pathological$_i$ may be nonpathological$_k$, or indeed true$_k$, $k > i$. *Pathologicality is not an intrinsic condition but a disposition to produce disease for certain semantical evaluations*—evaluations that in a context may or may not be implicated as appropriate. There is nothing wrong with deducing or asserting a pathological$_i$ sentence, as long as one's implicatures are respectable and as long as the sentence is true$_k$, for some relevant $k > i$. [Cf. step (b) in the examples above.] For the general case, interpretations of instances of our axioms are to be understood as carrying the next higher subscript.

On *Construction 1* (C1), the analogue of Tarski's, all and only sentences containing 'true$_k$' $k \geq i$, $i \geq 1$, are pathological$_i$. Pathological$_i$ sentences are not true$_i$. Instances of the truth$_i$ schema are asserted for all (and only) sentences that are not pathological$_i$. Construction 1 differs from Tarski's only in taking the application of 'true$_i$' to pathological$_k$, $k \geq i$, strings to be well-formed, and in appropriately conditionalizing the truth schema. The advantages of this difference are twofold. First, natural-language sentences that lead to paradox do not seem to be ungrammatical or in some cases even odd [cf. example (C)]. Second, the present construction allows us to give truth conditions at higher levels to predications that Tarski counts ungrammatical. This captures an intuition we want. The sentence in example (C) led to paradox when we applied the truth schema corresponding to the occurrence of 'true' in the sentence (call it 'true$_i$'), and was thus not true$_i$ (pathological$_i$). But

---

[20] Schemas like this are alluded to in passing by Parsons, op. cit., pp. 21, 28–29. A conceptually similar schema is mentioned by Kripke, op. cit., p. 81, although its intended interpretation is importantly different. Cf. also Buridan, op. cit., pp. 92, 192, 195.

since this is just what the sentence said, we want on reflection to call it true (true_{i+1}) after all.

Construction 1 agrees with Tarski's in immunizing pathological_i strings from the assignment of truth_i conditions via the truth_i schema. Self-referentially intended strings like 'This sentence is true_i' are not true_i—not because their truth_i conditions are not fulfilled (they have no truth_i conditions), but because they are pathological_i in not applying 'true_i' ('true' at the appropriate occurrence) derivatively.

Construction 1 rules pathological the sentences that are intuitively empty or lead to paradox. But to some (including myself) it seems too stringent. The intuition behind Construction 2 is that the results of logically valid_i inferences from true_i sentences that contain no predications of $\ulcorner\text{true}_k\urcorner$, $k \geqslant i$, are true_i. From the premise that 'All snow is white' is true_i, we could get that 'All snow is white or a is not true_i' is true_i (regardless of what 'a' denotes). Complex sentences that contain predications of $\ulcorner\text{true}_k\urcorner$, $k \geqslant i$, but whose truth_i or falsity_i is fixed by other components are nonpathological_i and can be given truth_i conditions.[21] On C2 a semantical evaluation of a sentence is derivative (nonpathological) only if the evaluation can be determined purely by reference to components of the sentence (or instances of its quantification) which either are non-semantical or are semantical predications with lower subscripts.

In what follows, I shall be assuming standard, first-order rules for well-formedness. What counts as a sequence may be determined by reference to any standard set theory. As for special vocabulary, let 'Sat_i' be the satisfaction predicate; 'P_i' represents 'is pathological_i (relative to an assignment)'; 'α' and 'α_1' range over sequences; 't' and 't_1' over terms; 'x' over variables; and 'Γ', 'β' and 'ϕ', over well-formed formulas (actually only 'ϕ' must be so construed). We shall indulge in the use of corners, understanding them to be convertible into a system of Gödel numbering or a concatenation theory.

Construction 2 (C2) is summed up in the following principles: First, the definition of 'is pathological_i':

(1) $P_i(\ulcorner\text{Sat}_k(t, t_1)\urcorner, \alpha)$, $P_i(\ulcorner P_j(t, t_1)\urcorner, \alpha)$ $k, j \geqslant i$

(2)   If $P_i(\Gamma, \alpha)$, then $P_i(\ulcorner \sim \Gamma\urcorner, \alpha)$

---

[21] The idea behind Construction 2 is very much like the intuition behind S. C. Kleene's strong tables for three-valued logic. Construction 1 is roughly analogous to the weak tables. See his *Introduction to Metamathematics* (Princeton, N.J.: Van Nostrand, 1952), pp. 332 ff. No doubt, other constructions are worth thinking through.

(3)   If $[P_i(\beta, \alpha) \vee P_i(\Gamma, \alpha)]$ and $[\sim\text{Sat}_i(\alpha, \ulcorner\sim\beta\urcorner) \wedge \sim\text{Sat}_i(\alpha, \Gamma)]$,
        then $P_i(\ulcorner\beta \supset \Gamma\urcorner, \alpha)$

(4)   If $(\exists\alpha_1)P_i(\phi, \alpha_1)$ and $\sim(\exists\alpha_1)(\alpha_1 \overset{x}{\approx} \alpha \wedge \text{Sat}_i(\alpha_1, \ulcorner\sim\phi\urcorner))$,
        then $P_i(\ulcorner(x)\phi\urcorner, \alpha)$

where '$\alpha_1 \overset{x}{\approx} \alpha$' means '$\alpha_1$ differs from α at most in its assignment to variable x'. The rules for determining pathologicality of conjunctions, disjunctions, biconditionals, and existential quantifications can easily be developed by reference to these.

(5)   An expression is pathological_i relative to an assignment only if it is so by (1)–(4) or by their analogues for other logical constants.

One point worth noting about this definition is that 'P_i' is relativized to a sequence even though this relativization plays no role in the basis clause (1). 'P_i' is defined partly in terms of 'Sat_i', which does not enter until clause (3). What I want to allow for is that an open sentence like '(x is not a mathematical sentence $\supset \text{Sat}_2(\alpha, x)$)' may be true_2 (or true_1) relative to one assignment to 'x' (e.g., "$2 + 2 = 4$"), but pathological_2 relative to another (e.g., "Dogs are mammals").

Now we connect 'P_i' and 'Sat_i', counting pathological_i sentences untrue_i:

(6) $P_i(\beta, \alpha) \supset \sim\text{Sat}_i(\alpha, \beta)$

We could, if we wished, *define* 'P_i' in terms of a sentence and its negation both being untrue_i.

We then relativize the semantical rules for the logical constants in the light of (1)–(6), arriving at a recursive characterization of truth:

(7) $\sim P_i(\ulcorner\sim\beta\urcorner, \alpha) \supset . \text{Sat}_i(\alpha, \ulcorner\sim\beta\urcorner) \equiv \sim\text{Sat}_i(\alpha, \beta)$

(8) $\sim P_i(\ulcorner\beta \supset \Gamma\urcorner, \alpha) \supset : \text{Sat}_i(\alpha, \ulcorner\beta \supset \Gamma\urcorner) \equiv . \text{Sat}_i(\alpha, \beta)$
        $\supset \text{Sat}_i(\alpha, \Gamma)$

(9) $\sim P_i(\ulcorner(x)\phi\urcorner, \alpha) \supset . \text{Sat}_i(\alpha, \ulcorner(x)\phi\urcorner \equiv (\alpha_1)(\alpha_1 \overset{x}{\approx} \alpha$
        $\supset \text{Sat}_i(\alpha_1, \phi))$

Similarly, for the other standard logical constants. We also note this principle, which has an analogue for 'P_j':

(10) $\sim P_i(\ulcorner\text{Sat}_j(t, t_1)\urcorner, \alpha) \supset . \text{Sat}_i(\alpha, \ulcorner\text{Sat}_j(t, t_1)\urcorner)$
        $\equiv \text{Sat}_j(\alpha(t), \alpha(t_1))$

where '$\alpha(t)$' means 'the assignment of α to t'.

The restricted truth schemas are

(T) $\sim P_i(S, \alpha) \supset . \operatorname{Sat}_i(\alpha, S) \equiv P$

where 'S' stands for the name of any well-formed sentence and 'P' stands for the sentence itself.

On both C1 and C2, truth is cumulative: a sentence that is true$_i$ is true$_k$, $k > i$:

(11) $\operatorname{Sat}_i(\alpha, \Gamma) \supset \operatorname{Sat}_k(\alpha, \Gamma) \qquad k > i$

From (11) it follows that a sentence not true$_k$, is not true$_{k-1}$, $k \geqslant 2$. As noted earlier, however, it is crucial that there be sentences not true$_i$, but true$_{i+1}$.

Iteration (on both C1 and C2) is appropriately expressed by ascending subscripts:

(12) $\operatorname{Sat}_i(\alpha, \Gamma) \supset \operatorname{Sat}_k(\alpha_1, \ulcorner \operatorname{Sat}_i(t, t_1) \urcorner), k > i; \alpha_1(t) = \alpha;$
$$\alpha_1(t_1) = \Gamma$$

A construction still more liberal in its certifications of non-pathologicality is possible. The intuition behind *Construction 3* (C3) is that if a sentence is true$_i$, then not only logically valid$_i$ inferences from it, but claims that it is true$_i$ are true$_i$.[22]

Thus we take all nonsemantical true$_1$ sentences, add sentences logically derivable from them; add all sentences that say that these sentences are true$_1$ (and that their negations are not true$_1$); then add sentences logically derivable from them; add all sentences that say that these are true$_1$ (and that their negations are not true$_1$); and so on. Then do the same for true$_2$, beginning with all true$_2$ sentences that either are nonsemantical *or contain only 'true$_1$'*. And so on. C3 differs from C2 in that it does not force all iteration to ascend a level. On C2 '"Snow is white" is true$_i$' is pathological$_i$ though true$_{i+1}$. On C3 the same sentence is true$_i$ as well as true$_{i+1}$.

The guiding idea of C3 is close to the view that sound semantical evaluation should be grounded.[23] The important difference is that C3 does not require that nonpathological semantical evaluations be grounded in nonsemantical soil; they may be *rooted* either in

---

[22] Substantially this idea was discovered independently by John Ruttenberg in my seminar on the paradoxes.

[23] For discussions of the notion of grounding, see Herzberger, "Paradoxes of Grounding in Semantics", *The Journal of Philosophy*, LXVII, 6 (Mar. 26, 1970): 145–67; and Kripke, op. cit.

nonsemantical statements or in lower levels of semantical evaluations. This departure is needed to account for intuitions about the move from (*b*) to (*c*) in our examples of section II.

C3 differs in its axioms from C2 essentially in that it weakens (1) and strengthens (12). But C3 is more perspicuously expressed in terms of a formula's being *rooted*, understanding pathologicality to consist in rootlessness. Letting 'R_i' mean 'is rooted$_i$ (relative to an assignment)', C3 is stated as follows:

(0) $R_i(\Gamma, \alpha)$, where the largest subscript in $\Gamma < i$.
(We understand that if there are no subscripts in Γ, 'the largest subscript in Γ' will denote 0.)

(1′)   If $R_i(\Gamma, \alpha)$, then $R_i(\ulcorner \operatorname{Sat}_i(t, t_1) \urcorner, \alpha_1)$ and
$R_i(\ulcorner R_i(t, t_1) \urcorner, \alpha_1)$, where $\alpha_1(t) = \alpha$ and $\alpha_1(t_1) = \Gamma$

(2′)   If $R_i(\Gamma, \alpha)$, then $R_i(\ulcorner \sim \Gamma \urcorner, \alpha)$

(3′)   If $[R_i(\beta, \alpha) \wedge R_i(\Gamma, \alpha)] \vee [\operatorname{Sat}_i(\alpha, \ulcorner \sim \beta \urcorner) \vee \operatorname{Sat}_i(\alpha, \Gamma)]$,
then $R_i(\ulcorner \beta \supset \Gamma \urcorner, \alpha)$

(4′)   If $(\alpha_1)R_i(\phi, \alpha_1) \vee (\exists \alpha_1)(\alpha_1 \overset{x}{\approx} \alpha \wedge \operatorname{Sat}_i(\alpha_1, \ulcorner \sim \phi \urcorner))$,
then $R_i(\ulcorner (x)\phi \urcorner, \alpha)$.

Rules for other logical constants are analogous.

(5′)   An expression is rooted$_i$ relative to an assignment only if it is so by (0′)–(4′) or their analogues.

(6′) $\sim R_i(\Gamma, \alpha) \supset \sim \operatorname{Sat}_i(\alpha, \Gamma)$

$\ulcorner P_i \urcorner$ is defined as $\ulcorner \sim R_i \urcorner$, and (12) is strengthened to

(12′) $\operatorname{Sat}_i(a, \Gamma) \supset \operatorname{Sat}_k(\alpha_1, \ulcorner \operatorname{Sat}_i(t, t_i) \urcorner)$
$$k \geqslant i; \alpha_1(t) = \alpha; \alpha_1(t_1) = \Gamma$$

Otherwise, the axioms of C2 carry over to C3, with the understanding that $\ulcorner P_i \urcorner$ changes its meaning according to the new definition.

It is possible to liberalize (1′) still further by changing the subscripts on 'Sat' and 'R' as they occur within the corners to j and k respectively, $j, k \geqslant i$. This allows as rooted$_i$ "loops" like

$$\text{'Tr}_i(\text{``Tr}_{i+3}(\text{`}2 + 2 = 4\text{')''})\text{'}$$

C3 is probably closest to intuition. But the key to the choice between C2 and C3 is iteration. If one could motivate the hierarchy better than I now know how to for *normal* cases of iteration ('"2 + 2 = 4' is true" is true), then C2 would become more attractive. And its ruling that "'2 + 2 = 4' is true," is pathologically not true$_i$

(though $true_{i+1}$) would take on appeal. I shall discuss differences among the constructions more concretely in section v.

On all three constructions, the law of excluded middle $\ulcorner P \vee \sim P \urcorner$ is valid (where validity attributions are subscripted in a manner suitable to the substitutions for 'P'). And all closed sentences are (indeed, everything is) $true_i$ or not $true_i$ for any i. We do relinquish the idea that every closed (or maximally interpreted) sentence or its negation is $true_i$. But we have direct intuitive evidence for this. Neither (α), "It is not the case that (α) is $true_i$," nor its negation should be counted $true_i$. Either (α) or its negation [namely, (α)] is, however, $true_k$, $k > i$.

What is the justification for making the relation between indexical uses linear, and the subscripts numerical? Having gone through the reasoning that leads to counting a $pathological_i$ sentence $true_k$, we can get ourselves into hot water again by adding, perversely, "But this very sentence isn't". We may regard ourselves as having intentionally and anaphorically taken over the context of use for '$true_k$'. To evaluate our perverse afterthought, we need a new context. So there is no limit on the number of different contextual applications of 'true' that might be required. Self-referential circles, like that in example (A) require that the relationship among markers of the contexts be transitive and asymmetric. Since sentences that do not contain semantical predicates (or other predicates of propositions, like 'knows', 'believes', or 'is necessary'—cf. note 3) do not produce paradoxes, it is natural to think of semantical predicate occurrences that apply to these sentences as having the lowest subscript.

To establish a linear relation, there remains only the requirement of comparability—that, for any two contexts of use for 'true', either the occurrences of 'true' have the same extension, or one of the occurrences is capable of evaluating or rationalizing the other. There is no compelling reason for this requirement in interpreting actual usage: does 'true' in example (B) have a higher or lower subscript than 'true' in example (C)? (One *could* relativize the hierarchy to a context, broadly conceived.) A natural consideration, however, leads to accepting the requirement. Any occurrence of a predicate can be assigned the lowest subscript compatible with the principles of material interpretation that we shall set out in section v. Assuming that the requirement is compatible with usage, I shall accept it as a means of rendering the formal model simpler.

What of the univocality criticism of Tarski? (Cf. section i.) Unlike Tarski, we do not interpret our systems as involving constant truth predicates.

In natural language there is a single indexical predicate. We represent this predicate by the schematic predicate expression $\ulcorner true_i \urcorner$. This expression may in particular contexts be filled out by any of an unlimited number of numerical subscripts. Any one of the resulting predicates (formally, there are infinitely many) may represent a particular occurrence of 'true' in a context in which its application is fixed. Thus numerals substituted for 'i' mark not new predicate constants, but contextual applications of the indexical 'true'. We have a general method for using this predicate. The existence of this method, which is represented in the formal principles given above and the material principles discussed in section v, provides considerable substance to the notion that 'true' has a single meaning.[24] On the other hand, the view that 'true' has a single extension is in conflict with intuitions about the Strengthened Liar [the moves from (*b*) to (*c*) in our examples].

A point often offered in favor of appealing to a "global" truth predicate is our ability to say such things as "All sentences are either true or not" or "God is omniscient". Such statements might be taken as asserted within a particular context (governed by a particular subscript). But one feels that they have broader import. These statements should be seen in the same light as principles, (0)–(12'). They are schematic generalizations. In the formal principles, the subscripts marking contexts of use stand open, ready to be filled in as the occasions arise. Similarly with our global English statements, including many statements in this paper. The first statement above, for example, should be formalized: $(s)(Tr_i(s) \vee \sim Tr_i(s))$. When we judge the schematic statement itself to be true, we make an equally schematic statement with the context for our evaluation schematically fixed as that of $\ulcorner true_{i+1} \urcorner$.

---

[24] I am ignoring the flexibility of 'true' as applied to different truth vehicles. As applied to sentences, two other contextual parameters enter in. The term must be relativized to a person and time to handle its application to nonsemantical, indexical sentences. And it must be seen to have another contextual parameter (not, I think, to be handled in tems of relativization to a language) which accounts for the fact that the same sentence may have different truth evaluations in different languages. Intralinguistic ambiguity is handled by interpreting the formalization of 'true' as applying to formal representations of the ambiguous surface sentence.

The indexical-schematic character of semantical predicates cannot be formally obviated by adding an argument place—relativizing them to a language, a level, a context, or a viewpoint. For quantification into the argument place will provide an open sentence just as subject to paradox as the "naive" truth-predicate formalization.[25]

Attempts to produce a "Super Liar" parasitic on our symbolism tend to betray a misunderstanding of the point of our account. For example, one might suggest a sentence like (a), '(a) is not true at any level'. But this is not an English reading of any sentence in our formalization. Our theory is a theory of 'true', not 'true at a level'. From our viewpoint, the latter phrase represents a misguided attempt to quantify out the indexical character of 'true'; it has some of the incongruity of 'here at some place'. No relativization will "deindexicalize" 'true'. Even in such English phrases as 'true at a level', the indexes occur implicitly on 'true'.

When we are given a semantical theory for nonsemantical indexical sentences, we relativize the semantical predicate to the context so as to generalize over all possible uses of the relevant indexical expressions. But insofar as we regard a semantical theory as a theory of the semantical predicate itself, there is no higher ground in which to absorb the indexical element. Theories of truth are in this sense models of or idealized directions for the use of the truth predicate. Axiom schemas like (7) are schematized directions for making statements whose component extensions are contextually fixed. The concept of truth cannot be defined or adequately represented in non-indexical terms. The indexical character of the language must be represented schematically.

<center>V</center>

The chief question about the application of the formal structure is how the subscripts are established in context. In our discussion of the linear structure of the hierarchy we suggested that the relevant subscript be the lowest subscript compatible with certain material

[25] Cf. Bertrand Russell, "Mathematical Logic as Based on the Theory of Types", in *Logic and Knowledge*, Robert C. Marsh, ed. (New York: Capricorn, 1971), pp. 64–9. See also Parsons, op. cit., note 13.

principles of interpretation. To begin with, natural-language statements are to be attributed no more pathologicality than other relevant considerations (specified below) dictate. More generally, subscripts on 'true' are assigned *ceteris paribus* so as to maximize the interpreter's ability to give a sentence truth conditions by way of a truth schema. We shall call this the *Principle of Verity*.

This principle is analogous to Quine's Principle of Charity in that it forms a general constraint on linguistic interpretation. But whereas Charity is motivated by an attempt to maximize the speaker's rationality, Verity applies in cases where the speaker's rationality is not at issue. If paradox is to be avoided, the subscript on a truth predicate in a quantified sentence of the form $\ulcorner(x)(Ax \supset \sim \mathrm{Tr}_i x)\urcorner$ must sometimes be higher than the subscript on truth predicates in sentences that satisfy A. For example, if someone said, "Everything Descartes said that does not concern mechanics was true", the subscript on 'true' would be high enough to interpret satisfactorily or give truth conditions to everything Descartes said that did not concern mechanics. The speaker and interpreter may not know what these subscripts are (even supposing they attached subscripts quite self-consciously) or be any less rational for their ignorance. The subscript (even if it remains schematic) is fixed by the context as the lowest that fits the interpreter's interpretative purpose.

The pragmatic justification for *Verity* should be pretty obvious. It excuses us in ordinary discourse from worrying about paradox, or semantical pathology generally, unless there is pressing reason to do so. Thus the extension of 'true' is a product not merely of the intentions of the speaker or hearer, but also of facts about the context of use and general conventions about the language. In this, it is roughly similar to the indexicality involved in a sign reading '(you) slow down'.

In the usual case, Verity will ensure healthy semantical statements. For example, if Construction 2 is preferred, normal iteration in surface English will be appropriately represented with ascending subscripts. How then does paradox arise? Sometimes the conditions laid down in a quantification or definite description (A in $\ulcorner(x)(Ax \supset \sim \mathrm{Tr}_i x)\urcorner$) will be clearly satisfied by the statement itself. If there is nothing in the subject's intentions, or in the context, that would warrant restricting the quantifier or descriptions, the statement is vulnerable to trouble. The student in example (C) unwittingly lands

himself in difficulty. Occasionally, the subject's own intentions force the issue—sometimes perversely, as with the original Epimenides form of the Liar, sometimes constructively, as in the argument for Tarski's theorem. The Principle of Verity then is prevented from sanitizing all discourse by standard conventions in interpreting the rest of a subject's expressions.

A less important principle is that of *Justice*. One should not give one statement truth conditions instead of another without some reason. In (A), for example, there is no evident reason for treating Plato's statement any differently from Aristotle's. Although we are logically forced only to deny that both statements can be true, we should *ceteris paribus* assign both predicates the same subscript (1) and count both pathological.

Let us see how these principles operate to solve a problem that has been raised against Tarski's treatment. Suppose Dean says:

(i)   All Nixon's utterances about Watergate are untrue

and Nixon asserts

(ii)   Everything Dean utters about Watergate is untrue.

Each wishes to include the other's assertion within the scope of his own assertion. To ensure Justice, each person's truth predicate should be assigned the same subscript, i. To ensure Verity, we assume i is high enough to interpret any statement by Dean or Nixon other than (i) or (ii). I shall discuss the example on Constructions 2 and 3, which handle it deftly. On C2, in evaluating (i) and (ii) *we* use 'true$_{i+1}$', since on this approach sound semantical evaluation will be forced to a higher level. On C3 we must use 'true$_i$'. I shall place C3's reasoning in parentheses.

Suppose Dean has uttered at least one truth$_i$ about Watergate. It follows from the semantical rules for the quantifier [cf. (4), (5), (4′), (12′), (9)] that Nixon's assertion (ii) is nonpathological$_{i+1}$ and not true$_{i+1}$ (also nonpathological$_i$ and not true$_i$ on C3). If none of Nixon's other Watergate utterances besides (ii) are true$_i$, then since (ii) itself is not true$_i$ [since it is pathological$_i$ by (1)–(6) on C2; since its truth$_i$ conditions are not fulfilled on C3], Dean's (i) is true$_{i+1}$ (also true$_i$ on C3). On the other hand, if Nixon eked out at least one true$_i$ statement, then Dean's (i) is not true$_{i+1}$ (also not true$_i$ on C3). By erasing the subscripts and ignoring the parenthetical remarks in the previous four sentences, we have a piece of reasoning that is intuitive. Our theory accounts for the reasoning.[26]

The difference between C2 and C3 can be further elucidated by the following example.[27] Suppose Nixon says

(iii)   Mitchell is innocent and (iv) is not true

and Dean says

(iv)   (iii) is not true.

Let the occurrences of 'true' in (iii) and (iv) be marked by 'true$_i$'. Since (iii)'s first conjunct is not true$_i$, (iii) has truth$_i$ conditions and is not true$_i$ on both C2 and C3; so (iv) is true$_{i+1}$ (true$_i$ and true$_{i+1}$ for C3). Now there is a tendency for us to reason

(v)   Since (iv) is true and the second conjunct of (iii) says it's not, the second conjunct cannot be true.

The two constructions handle the case differently. C2 can accept (v) only if it is understood to involve a potential equivocation. (iv), by the preceding reasoning, is true$_{i+1}$. The second conjunct of (iii), interpreted not as it is in the context of Nixon's statement but as a denial of ours ['(iv) is not true$_{i+1}$'], is not true—not true$_{i+2}$. By contrast, the second conjunct of (iii) as it is interpreted in *Nixon's* statement ('(iv) is not true$_i$') is in no position to evaluate Dean's (iv) (according to C2) since it is not appropriately derivative. Dean's (iv) is trivially not true$_i$ since it lacks truth$_i$ conditions. So the second conjunct of (iii) interpreted as it is in Nixon's own statement is trivially true$_{i+1}$—not because Dean's (iv) is false$_i$ as Nixon would like us to imagine, but because it lacks truth$_i$ conditions. C2 suspends Nixon's right to evaluate Dean's (iv) because of the mutually reflexive situation. But it accounts for our ability to adjudicate the situation. Interpreted as a judgment from our point of view, (v) is justified.

Construction C3 treats the matter in a more straightforward and, I think, more natural way. Since (iv) is true$_i$ and rooted$_i$, and (iii)'s second conjunct says it's not true$_i$, that second conjunct is rooted$_i$ and not true$_i$. No fundamental distinction is drawn between our

---

[26] Kripke, op. cit., pp. 59–60, uses the example to show that Tarski's method of fixing the levels (applied literally to natural language) would counter-intuitively pronounce at least one of the statements ill-formed or nonsensical.

[27] I owe the example to Nathan Salmon, who had a different purpose in mind.

evaluation and Nixon's. Our intuitions about what the protagonists in such semantic entanglements can or cannot do are perhaps not clear-cut. Still, in the absence of reasons to the contrary, C3 is probably to be preferred.

VI

Let us survey the dividends of our account. We make no change in classical logic, no general restriction on quantification, no unintuitive postulations of ungrammaticality or meaninglessness. Our restrictions on the applications of 'true' are directly motivated by intuitive considerations. The theory provides a basis for explicating the univocality of 'true'. It gives weight to intuitions both about the "global" character of some uses of 'true' and about the context-dependent character of others. And it accounts for rather than merely obstructs paradoxical reasoning.

A bonus is that the account places no unnatural restrictions on translation of semantic discourse between natural languages. One of Tarski's characterizations of universality of a natural language is that any word in another language can be translated. Some writers seeking to apply Tarski's theory have argued that natural languages are not universal in this sense, holding for example that our predicate 'true in Urdu' cannot be translated into Urdu.[28]

The reasoning seems to go somewhat as follows. If Tarski's theory is to be applied to natural language, one must take a semantic system like his (including semantic postulates) as standing for or representing a natural language. A truth predicate in a natural language (e.g., 'true in English' or 'true in Urdu') should be represented by a predicate constant, with a fixed extension (e.g., all the true sentences of English) determined by the predicate's form and meaning. If Tarski's theory is to be applied, this constant must be governed by the usual semantic postulates. But, by Tarski's theorem, any such predicate for evaluating all the sentences of a

[28] Tarski, op. cit., 164; Herzberger, "Paradoxes of Grounding", op. cit., p. 167; Donald Davidson, "Truth and Meaning", *Synthese*, XVII, 3 (September 1967): 304–23, pp. 313/14. I have no desire to get into the somewhat vague question of whether artificial languages with non-indexical semantic predicates could always be translated into English. Of course, I think natural languages are not universal or closed, under Tarski's more precise characterization in terms of the capacity for self-reference and an unrestrictedly applicable truth schema with a truth predicate with constant extension. Cf. note 2, above.

semantic system cannot be introduced and used in the semantic system (with the usual semantic postulates) on pain of contradiction. So if Tarski's theory is to be applied, a truth predicate like 'true in English' cannot be allowed or cannot occur in English itself. Roughly this argument has played a role in criticisms (e.g., Tarski's criticism) as well as defenses of applying Tarski's theory to natural language, the difference in opinion focusing on whether the conclusion is a *reductio* of the initial if-clause.

Our account rejects the two initial assumptions of the argument. First, as noted at the outset, natural languages are not the sort of thing that can be inconsistent. One cannot assume that Tarski's "object language", "metalanguage" terminology can be cashed out in ordinary "language" language. So consistency restrictions on formal definability or formal introduction in a theory have little to do with conditions on translatability between natural languages. In the second place, Tarski's results bear on the definability or introducibility of predicate constants with an intuitively fixed extension. But the predicate 'true' (or 'true in Urdu') is not fixed apart from contexts of use. Tarski's results do bear on what extensions the predicate can have in given contexts. But they cannot prevent the occurrence or use of such a predicate even within consistent (context-dependent) formal systems, much less in natural languages.

On our view, 'true in Urdu' ('true$_i$ in Urdu') translates into Urdu without difficulty. The context-dependence and "implicit" subscripts are no less present when the predicate is used in English than when its analogue is used in Urdu. And this feature should be preserved under translation. The principles for establishing the level of a subscript on 'true' are not motivated purely by a desire to avoid contradiction. They are designed to capture the derivative feature of semantic evaluation in natural discourse.

Our reflections have suggested two general aspects of the use of indexical semantic predicates like 'true'. One is that their application is *derivative*. Their correct application is to statements which can be formulated and which have sense and reference, independently of the application. As a consequence, no statement can sit in semantic judgment on itself. Russell's vicious-circle principle and Tarski's appeal to a metalanguage were attempts to elucidate this important aspect of semantic notions. Truth-value gaps articulate it in their own way. Redundancy theories (e.g., Ramsey's and Strawson's) represent an extreme emphasis on it. A second aspect of

our use of 'true' is that its applications are *evaluative*. In using the term we scrutinize sentences or statements to determine whether they are factually satisfactory, or, more loosely, whether things are as they are represented. Tarski's target biconditionals and his accompanying semantical analysis constituted a brilliant illumination of the structure of this evaluative use. Aristotle's well-worn dictum and traditional correspondence theories (e.g., the early Wittgenstein's and Austin's), for better or worse, were inspired by it.

The approaches to the paradoxes that we have criticized treat the derivative feature of semantical predicates as a fixed or absolute limitation on their evaluative use. Such approaches do not work because reflection on the proposed solution (in the Strengthened Liar) produces a new evaluation which cannot be expressed in terms of the solution or which is incompatible with it. The intuitive staying power of the evaluative use of the semantical predicates has been seriously underestimated in most post-Tarskian discussions: we have evaluative intuitions even in pathological cases. Semantical paradox issues from counterclaims between the derivative and evaluative aspects of semantical predicates. Our theory describes laws that resolve the conflict, while attempting to do justice to both claimants.

Postscript to "Semantical Paradox", 1982

"Semantical Paradox" is guided by two ideals. One is that it is possible to accommodate specific judgments about truth that arise in the course of reasoning that leads to paradox. Specific judgments are to be distinguished from those that attach to generalizations, principles, or schemas about truth. The distinction has, of course, a fuzzy borderline and should not be relied upon heavily; but I think it useful. The other ideal is that it is possible to identify in a semantical and pragmatic theory actual aspects of language or thought whose neglect yields the paradoxes. There are many ways to "block" the paradoxes. Any number of devices, provisions, or systems can be invoked to do so. A few of these have independent mathematical interest. Yet most ignore specific, widely shared judgments or propose theories whose distinctions are *ad hoc*, at least considered as accounts of actual usage, and which do not cohere well with the rest of linguistic theory. Our aim is to dissolve the paradoxes by accounting for specific judgments by means of a theory of language

that does not require us to make implausible claims about the linguistic or pragmatic properties of the discourse, and that is motivated as directly as possible by those judgments.

These ideals are vague. And it is certainly not clear that they will determine a unique theory. Nevertheless, they have seemed to me to have considerable restrictive force, when applied seriously.

Except for the alteration of principle (0) (noted above), the basic theory is not much changed since its original publication. One refinement is a canonical ordering governing application of the pragmatic principles. The proper ordering seems to be: *Justice, Verity, Minimalization* (perhaps better labeled '*Beauty*'). This point and several special applications are discussed in "The Liar Paradox: Tangles and Chains", *Philosophical Studies* (1981).

Other possible refinements through application concern the variety of other notions that (arguably) are paradox-producing and that are not strictly semantical: class membership, necessity, knowledge, belief, acceptance, fearing, wanting, saying, promising, ordering, and so forth. I intend to use the theory to compare and contrast semantical notions and cognitive notions like belief and occurrent acceptance.

There is, of course, considerable room for technical refinement and development. Making explicit provisions for extending the constructions into the transfinite is of particular technical importance. The lack of such provisions in 'Semantical Paradox', together with some remarks that were intended as merely illustrative, have misled some into thinking that I intended to restrict the subscripts on semantical predicates to finite levels. In fact, there are no such restrictions on the subscripts in our statement of the formal principles. I believed and still believe that no such restrictions are appropriate. I was persuaded of this by Bill Hart in the course of writing the paper, and my views have since been enriched by work of Charles Parsons. Provisions must be made for the subscripts to range over transfinite ordinals and for the associated limit levels. I did not confront this issue in 'Semantical Paradox' because I realized that it raised substantial mathematical and conceptual difficulties and because I believed (and believe) that their solution would not profoundly affect the basic approach that I proposed in the article. I should, however, have made my view on this more explicit.

A further sort of refinement I envision concerns the philosophical interpretation of the theory. The theory is committed to there being

two uses of 'true' in natural language: indexical uses and schematic uses. A predicate is *indexical* on an occasion of use if and only if it has a definite, fixed extension (or extensional application) on that occasion that depends not only on the contextually appropriate conventional meaning of the predicate, but further on the immediate context of its use. A predicate is *schematic* on an occasion of use if and only if it lacks a definite extension on that occasion, but through its conventional use on that occasion provides general systematic constraints on the extension(s) of the same predicate (or importantly related ones) on other occasions of use. The specific examples discussed in "Semantical Paradox" largely concern indexical uses. (There is no claim, incidentally, that the theory itself uses indexicals.) The formal principles are stated, as they must be, schematically. But these formal principles, and the pragmatic ones, apply to both indexical uses and schematic uses. This remark should forestall the confused criticism, which I have heard twice, that the basic formal principles are, according to the theory, neither true nor false. Such principles are true. In saying so we are using 'true' schematically.

Use of schematic principles in mathematical logic is common and well-established. I believe that it is primitive and not in general eliminable. We can express the "generality" intended by the subscript quantificationally. But in so doing, we invoke a metalinguistic formula and a further semantical predicate. This predicate will itself be schematic. Thus we can say schematically: $\ulcorner\phi$ is true$_i\urcorner$ (suppose 'ϕ' is a name of a formula). We can express the intended "generality" in the form of a quantification:

For any ordinal number i $\ulcorner\phi$ is true$_i\urcorner$ is true.

But this latter occurrence of 'true' will also be schematic. There is no deschematizing the schema.

In view of its mathematical entrenchment and usefulness, schematic usage can hardly be seen as mysterious in the sense of 'suspect'. On the other hand, there is considerable room for improved philosophical understanding of it. The distinction between indexical and schematic uses connects with some of the most profound and difficult questions in interpreting foundational theories in mathematics, both type theories and set theories. One avenue that promises to deepen understanding is the comparison between schematic uses of 'true' and discourse about classes (as distinguished from sets) in set theory. There are several formal and intuitive parallels between indexical (extension-fixing) uses of 'true' and talk about sets, on one hand, and between schematic uses and talk about classes, on the other.

The circle of related notions may be wider. I think that the attempts by Reinhardt and Parsons to say something about the class–set distinction by reference to modal principles may be useful in illumining the indexical–schematic distinction as applied to 'true of'. (W. N. Reinhardt, "Remarks on Reflection Principles, Large Cardinals, and Elementary Embeddings", in *Axiomatic Set Theory* (*Proceedings of Symposia in Pure Mathematics*, vol. 13, 1974), part II, pp. 189–206; Charles Parsons, "What is the Iterative Conception of Set?" in *Logic, Foundations of Mathematics, and Computability Theory* (D. Reidel, Dordrecht, 1977).)

On the other hand, as Parsons notes, the analogy between modal notions and class-like notions is limited by the fact that the set–class distinction infects the interpretation of the modal language itself. Moreover, if (as I believe) the fundamental notions of necessity are expressed via predicates of sentential or propositional entities, rather than as intensional operators, the modal notions themselves will be expressed in indexical–schematic language. For, as is well-known, modal paradoxes analogous to the Liar emerge in a language in which modality is expressed as predication of sentences (or propositions with something like sentential structure). I believe that there is no transcending the indexical–schematic distinction—or reducing it to other terms, such as modal terms. But understanding it can benefit from structural and intuitive analogies to other conceptual systems.