TYLER BURGE

BURIDAN AND EPISTEMIC PARADOX*

(Received 4 June, 1977)

I

Near the end of *Sophisms on Meaning and Truth* Jean Buridan presents a group of epistemic paradoxes that bear a close relation to modern formulations of the Hangman Paradox.[1] What is interesting about them is that they seem to teach a lesson not immediately evident in the latter-day puzzle. My purpose here is to analyze and amplify these paradoxes, articulate the lesson, and criticize Buridan's solution.

In Sophism 13 of Chapter VIII Buridan supposes that the following proposition is written on the wall:

Socrates knows the proposition written on the wall to be doubted by him.[2] Socrates reads it, thinks it through and is unsure (doubts) whether or not it is true. Further, he knows that he doubts it. Buridan asks whether or not the proposition is true.

Buridan gives two arguments for its truth, and one against. (A) By hypothesis, Socrates doubts the proposition and knows he does. But this is what the proposition says. So it is true. The second positive argument is (B) that anyone who uttered another token of the same sentence would be regarded as speaking the truth. But (to amplify Buridan's brief remarks) such a token has exactly the same component references and extensions as the token on the wall. So the two have the same truth value. Against the truth of the proposition, Buridan argues (C) that if it were true, Socrates would *per impossible* know and doubt the same proposition. For by hypothesis, he doubts it. Yet if we only assume that he knows that *he knows that he doubts the proposition on the wall*, we can derive that he knows it. (For the sentence in italics itself represents the proposition on the wall.)

Buridan attempts to resolve his puzzle by attacking argument (C) and counting the proposition true. He begins by distinguishing between primary knowledge and remote knowledge. Primary knowledge is roughly assent to a

Philosophical Studies **34** (1978) 21–35. *All Rights Reserved*

true proposition with evidence and certainty. Remote knowledge is knowledge of the things signified (denoted) by terms of a proposition primarily known. For example, Jean knows John's favorite proposition remotely if Jean knows (in the primary sense) that John's favorite proposition is mathematical. (Buridan does not say explicitly that remote knowledge is propositional, but it does not seem unreasonable to assume that Jean remotely knows of John's favorite proposition that it is mathematical. On this interpretation, the distinction bears a rough parallel to oblique and transparent knowledge attributions.) Buridan claims that the (mental) proposition that Socrates knows in the primary way is different from the one on the wall – which he doubts – and that he knows the one on the wall only remotely: he knows *of it* that he doubts it. Thus, according to Buridan, the known and the doubted propositions are different, though they have the same truth value (they are both true.) If one asks how one could know the mental proposition (token) and doubt the written proposition (token), given their similarity, Buridan answers that the written one has the important property of self-referentiality, which the mental one lacks.

For present purposes I will not question Buridan's claim that the written token and the mental token are different propositions. We can grant the possibility of doubting one and believing the other and still see that the solution is so far unsatisfying. For the two tokens have the same component denotations and extensions; Buridan himself holds that they have the same truth value. And if Socrates were sufficiently reflective, he could come to the same conclusion.

In the last paragraph of his discussion of the sophism, Buridan responds to this sort of point. He postulates that Socrates is 'most wise', by which he means that Socrates can reason through the case in a theoretically ideal manner. From this Buridan concludes that *the original hypothesis* that Socrates doubts whether or not the written proposition is true, *must be withdrawn*. For if the written proposition were true, Socrates, being 'most wise', would know it to be true; and if it were false, he would know it to be false. Since Socrates would not doubt the proposition on the wall (in the sense of being uncertain), he could not know himself to doubt it. So assuming Socrates is most wise, Buridan concludes that the written proposition is false, and that Socrates would know that it is false. Socrates' own reasoning could follow argument (C). A distinction in attitude to the mental and the written proposition need not be invoked.

This resolution seems correct as far as it goes. But it does not confront an obvious strengthening of the original case that remains troublesome. If we replace 'doubts' by 'does not believe', paradox returns with a vengeance. Suppose that the following proposition were written on the wall:

Socrates knows that he does not believe the proposition on the wall. We assume that Socrates is 'most wise' and that he reads the writing, and reflects upon it. Let us abbreviate 'the proposition on the wall' by 'a'. 'K' will represent the relation of knowing between people and propositions; 'B' represents the analogous belief relation; 'S' denotes Socrates. We shall denote propositions by both cornerquoting and overlining expressions that are replicas of (or express) the relevant proposition. (We assume that these devices are equivalent to Gödel numbering or concatenation devices.) 'I' abbreviates 'logically or arithmetically implies'. Then the following assumptions are analogous to Buridan's:

$$(0) \qquad a = \ulcorner K(S, \overline{-B(S, a)})\urcorner .$$

The proposition on the wall is the proposition that Socrates knows he does not believe the proposition on the wall.

$$(1) \qquad K(S, \overline{-B(S, a)}) \rightarrow K(S, \ulcorner K(S, \overline{-B(S, a)})\urcorner).$$

If Socrates knows he does not believe a, then he knows that the knows he does not believe a.

$$(2) \qquad K(S, a) \rightarrow B(S, a).$$

If Socrates knows a, he believes a.

$$(3) \qquad K(S, \overline{-B(S, a)}) \leftrightarrow -B(S, a).$$

Socrates knows he does not believe a if and only if he does not believe a.

$$(4) \qquad K(S, \overline{(0) \,\&\, (1) \,\&\, (2) \,\&\, (3)}).$$

Socrates knows the preceding four principles.

$$(5) \qquad I(\overline{(0) \,\&\, (1) \,\&\, (2) \,\&\, (3)}, \ulcorner -K(S, \overline{-B(S, a)})\urcorner) \,\&\, K(S, \overline{(0) \,\&\, (1) \,\&\, (3)}) \colon \rightarrow K(S, \ulcorner -K(S, \overline{-B(S, a)})\urcorner).$$

If (i) principles (0)–(3) logically or arithmetically imply that Socrates does not know that he does not believe *a* and (ii) Socrates knows principles (0)–(3), then Socrates knows that he does not know that he does not believe *a*.

(0) expresses an empirical fact. (2) and the left-right direction of (3) are instances of basic principles about knowledge (roughly, that knowledge implies belief and truth). (1), the right-left direction of (3), and (4)–(5) express our assumptions about Socrates' reflection and reasoning powers. These principles are not fully generalizable, but they are plausible in many particular cases including the present one. We assume for the sake of argument (and until we discuss Buridan's solution) that in the light of Socrates' reflection and wisdom, differences among proposition tokens that have the same component denotations and extensions can be ignored. (In the present case, component meanings are the same as well.) Now if we add one more formula connecting knowledge and belief and expressing an assumption about Socrates' wisdom, we can derive a contradiction:

(6) $K(S, \ulcorner -K(S, \overline{-B(S, a)})\urcorner) \rightarrow -B(S, \ulcorner K(S, \overline{-B(S, a)})\urcorner)$.

Given that Socrates knows the negation of something, he does not believe that something – at least in this particular case.)

The contradiction is obtained as follows:

(a) $(3) \vdash K(S, \overline{-B(S, a)}) \rightarrow -B(S, a)$,
(b) $(0), (1), (2) \vdash K(S, \overline{-B, a)}) \rightarrow B(S, a)$,
(c) $(0), (1), (2), (3) \vdash -K(S, \overline{-B(S, a)})$.

By the fact that the relation of derivability can be represented within elementary syntax (hence within number theory):

(d) $\vdash \mathrm{I}(\overline{(0) \,\&\, (1) \,\&\, (2) \,\&\, (3)}, \ulcorner -K(S, \overline{-B(S, a)})\urcorner)$,
(e) $(5) \vdash K(S, \overline{(0) \,\&\, (1) \,\&\, (2) \,\&\, (3)}) \rightarrow K(S, \ulcorner -K(S, \overline{-B(S, a)})\urcorner)$,
(f) $(4), (5), \vdash K(S, \ulcorner -K(S, \overline{-B(S, a)})\urcorner)$,
(g) $(4), (5), (6) \vdash -B(S, \ulcorner K(S, \overline{-B(S, a)})\urcorner)$,
(h) $(0), (3), (4), (5), (6) \vdash K(S \overline{-B(S, a)})$.

Informally, the reasoning is as follows: (a) Given (3), if Socrates knows he does not believe *a*, then he does not believe *a*. (b) By (1), if Socrates knows he does not believe *a*, he knows this. But 'this' is just *a* (by (0)). By (2), if Socrates knows *a*, he believes *a*. So if Socrates knows he does not believe *a*, he does believe *a*. (c) Since the assumption that Socrates knows he does not believe *a* leads to contradictory conclusions, Socrates does *not* know he does not believe a. (d) The preceding shows that (0)–(3) logically (or arithmetically) imply that Socrates does not know that he does not believe a. (e)–(f) By (4) and (5), Socrates can reason through the case as well as we can. So he knows that he does not know that he does not believe a. So (g), given (6), he does not believe that he knows that he does not believe a. (h) Given (0), we can substitute in (g) to conclude, Socrates does not believe a. But then by (3), Socrates knows he does not believe a. (h) and (c) are contradictory.

The argument is a genuine paradox since all the premises have some plausibility, yet they lead to contradiction. I have developed the argument rigorously and in detail to bring out its paradoxical character. For often Buridan's own exposition suggests that he is dealing with mere sophisms in which the problem is merely to spot, without theoretical development, an equivocation or a false premise.

In the two sophisms following 13, Buridan delves more deeply into paradox. Since 14 and 15 are for theoretical purposes equivalent, I shall focus only on 14. Buridan postulates that Plato, who is 'most learned' (inferentially and theoretically ideal, but not empirically omniscient). By hypothesis, Plato is unsure whether or not Socrates is sitting, and is confronted by the following proposition written on the wall:

Socrates sits or the disjunction written on the wall is doubted by Plato. Letting a_1 be the relevant proposition, 'P' denote Plato, 'p' stand for 'Socrates sits', and 'D' represent the relation of doubting, the premises of the crucial part of Buridan's argument may be represented as follows:

(7) $\quad a_1 = \overline{p \vee D(P, a_1)}$,

a_1 is the proposition that either Socrates sits or Plato doubts a_1.

(8) $\quad D(P, a_1) \to K(P, \overline{D(P, a_1)})$.

If Plato doubts a_1, he knows he does.

(9) $\quad K(P, a_1) \to -\overline{D(P, a_1)}$.

If Plato knows a_1, he does not doubt a_1.

(10) $\quad K(P, D(P, a_1)) \to K(P, \overline{p \vee D(P, a_1)})$.

If Plato knows he doubts a_1, he knows the relevant disjunction.

(11) $\quad K(P, \overline{(7) \mathbin{\&} (8) \mathbin{\&} (9) \mathbin{\&} (10)})$.

Plato knows the previous four principles.

(12) $I(\overline{(7)\ \&\ (8)\ \&\ (9)\ \&\ (10)}, \overline{-D(P, a_1)})\ \&\ K(P, \overline{(7).\&(8)\&(9)\&(10)})$
$\rightarrow K(P, \overline{-D(P, a_1)})$.

If (i) $\overline{(7)\text{–}(10)}$ implies that Plato does not doubt a_1 and (ii) Plato knows those principles, then he knows he does not doubt a_1.

(13) $D(P, \bar{p})$.

Plato doubts whether Socrates sits.

(14) $K(P, \overline{-D(P, a_1)})\ \&\ \mathrm{D}(P, \bar{p})\ .\rightarrow D(P, \overline{p \vee D(P, a_1)})$.

If Plato knows the negation of one disjunct and doubts the other disjunct, he doubts the disjunction.

The argument to contradiction goes as follows:

(a) $(8), (10) \vdash D(P, a_1) \rightarrow K(P, \overline{p \vee D(P, a_1)})$,
(b) $(7), (8), (9), (10) \vdash D(P, a_1) \rightarrow -D(P, a_1)$,
(c) $(7), (8), (9), (10) \vdash -D(P, a_1)$,
(d) $I(\overline{(7)\ \&\ (8)\ \&\ (9)\ \&\ (10)}, \overline{-D(P, a_1)})$,
(e) $(11), (12) \vdash K(P, \overline{-D(P, a_1)})$,
(f) $(11), (12), (13), (14) \vdash D(P, \overline{p \vee D(P, a_1)})$,
(g) $(7), (11), (12), (13), (14) \vdash D(P, a_1)$,

(g) and (c) show that the premises are mutually contradictory.

Informally the reasoning goes as follows. (a) By (8), (10) and the transitivity of the conditional, if Plato doubts a_1, he knows that either Socrates sits or Plato doubts a_1. (b) Since this disjunction is a_1 (by (7)), it follows that if Plato doubts a_1, he knows a_1. But since knowing implies not doubting, (9), it follows that if Plato doubts a_1, he does not doubt a_1. So (c) by propositional calculus, he does not doubt a_1. Thus (d) (7)–(10) logically imply that Plato does not doubt a_1. (e) Plato can reason through the preceding as well as we can. So he knows he does not doubt a_1. (f) But given he knows the negation of one disjunct and doubts the other disjunct (by (13)) (doubts whether Socrates sits), we conclude by (14) that he doubts the whole disjunction – he is in doubt whether either Socrates sits or Plato doubts a_1. (g) But by (7), this is to doubt a_1. (c) and (g) are contradictory.

Buridan does not make the argument from (d) through (g) explicit. Nor does he explicitly state premises (11), (12), or (14), though each could be said to be encompassed by the assumption that Plato is most learned. But Buridan appears to have the argument in mind, at least sketchily. For he remarks, at greater length in Sophism 15 than Sophism 14, that the relevant proposition a_1 is indeed doubted by Plato. How he reconciles this remark with his solution will be the subject of Section III.

II

A simpler paradox can be distilled out of the previous ones. The following assumptions are mutually inconsistent:

$$(15) \qquad a_2 = \overline{-B(P, a_2)},$$

a_2 is the proposition that Plato does not believe a_2.

$$(16) \qquad -B(P, a_2) \rightarrow K(P, \overline{-B(P, a_2)}).$$

If Plato does not believe a_2, he knows he doesn't.

$$(17) \qquad K(P, a_2) \rightarrow B(P, a_2).$$

If Plato knows a_2, he believes a_2.

$$(18) \qquad B(P, a_2) \rightarrow K(P, \overline{B(P, a_2)}).$$

If Plato believes a_2, he knows he believes it.

$$(19) \qquad K(P, \overline{B(P, a_2)}) \rightarrow -B(P, \overline{-B(P, a_2)}).$$

If Plato knows he believes a_2, he does not believe that he doesn't believe a_2. (I leave it to the reader to derive the contradiction.)

The arguments that stem from (1)–(6), (7)–(14), and (15)–(19), are striking because of their independence of the specific remedies that suffice to cure similar pathologies. From Tarski's work, we know that the semantical paradoxes (the Liar, Grelling's, Richard's, Berry's) can be blocked by introducing a hierarchy of semantical predicates. Instances of a given truth schema of the form

(T) '__________' is true if and only if __________

are to be asserted (or proved as theorems) only for sentences substituted in the blanks that contain no semantical predicate of level greater than or equal to the level implicitly associated with 'true' as it occurs in (T).[3]

Ramification of the notion of truth can be mobilized to deal with certain indirect discourse and epistemic versions of the Liar paradox. For example, such ramification suffices to handle difficulties arising from the supposition that A is thinking that $2 + 2 = 4$, B is thinking that dogs are reptiles, and C is thinking that an odd number of the current thoughts by A, B, and C are true. Or, to take a commoner case, the Eubulidean version of the liar ('What I am now saying is not true') can be disarmed by focusing on the notion of truth.

A pair of papers by Kaplan–Montague and Montague have since shown that the notions of knowledge and necessity are sufficiently analogous to truth in their formal properties to lead to epistemic and modal paradoxes similar to the semantical ones, even if the term 'true' plays no role in the argument.[4] These paradoxes can be obstructed by treating 'knows' and 'is necessary' as analogous to 'is true' in involving a hierarchy.[5] Thus, on this approach, instances of such principles as

(K) If S knows ϕ, then ϕ,

(cf. premise (ii) in Note 4) are to be asserted only for substitutions for 'ϕ' containing no predicate 'knows' (or semantical or modal predicate) with a subscript greater than or equal to the subscript implicitly carried by the 'knows' in (K). Given some development, this approach is, I think, basically on the right track, and can be defended as giving a reasonable account of 'is true', 'knows' and so forth. (Cf. Note 4.)

A natural observation at this point is that all of these notions – truth, denotation, necessity, knowledge – are similar in being governed by reflexion principles of the sort illustrated by (T) and (K). Belief and doubt are not in general governed by such principles. From the fact that someone believes something, we cannot infer that something. More generally, ascriptions of belief and doubt do not in any simple way, involve a commitment to an evaluation in a way that ascriptions of truth, knowledge and necessity do.

These observations might well sustain the hope that belief, doubt and other propositional-attitude notions could be exempt from the hierarchical restrictions standardly placed on truth and extended by analogy to knowledge. Buridan's paradoxes undermine this hope. The standard hierarchical restrictions do not suffice to block any of the above arguments, if those restrictions are applied *solely* to 'knows' ('K') and 'true'.

In fact, one can eliminate knowledge and truth altogether and still produce paradox. If all occurrences of 'K' are deleted in favor of occurrences of 'B'

in (15)–(19) – with the analog of (17) dropped since it becomes a tautology –, we still have a set of plausible principles that lead to contradiction:[6]

$$
\begin{array}{ll}
(20) & a_2 = \overline{-B(P, a_2)}, \\
(21) & -B(P, a_2) \rightarrow B(P, \overline{-B(P, a_2)}), \\
(22) & B(P, a_2) \rightarrow B(P, \overline{B(P, a_2)}), \\
(23) & B(P, \overline{B(P, a_2)}) \rightarrow -B(P, \overline{-B(P, a_2)}),
\end{array}
$$

Thus from the hierarchical viewpoint, the hierarchy must be extended to cover ordinary propositional attitude notions that bear no simple relation to truth. Different levels of belief, doubt and the like must be distinguished to avoid the conundrums.

Of course, (21) and (22) are formally isomorphic to certain instances of the principle: if __________ then '__________' is true. And perhaps the plausibility of (21) and (22) depends on attributing to Socrates, in the particular case, an ability to *accurately* represent to himself through introspection of his own beliefs or representations. Our assumption that people sometimes reflect on their own beliefs and the limitations we place on the corrigibility of a person's beliefs about his own mental states are in a sense two aspects of a normative feature of psychological notions. So in some deep sense truth may be lurking at the bottom of the paradoxes after all. My present point is a formal one. Restrictions on semantical notions, or on notions like knowledge that bear simple entailment relations to semantical notions, do not suffice to prevent paradox.

What leads to this last paradox of belief is the self-reflexive character of the original proposition (20), an assumption about the believer's consistency or rationality (23), and two assumptions about the believer's selfconsciousness, (21) and (22). From the viewpoint of the hierarchical approach, the paradox elicits the fact that belief is a *derivative* notion in a way analogous to the ways in which truth and (on the iterative conception) set are derivative notions. A sentence like

(α) α is true

is empty or pathological because (one feels) nothing has been said that can be evaluated as true. Something independent of the evaluation must be expressed before semantical evaluation is possible. Similar remarks apply to knowledge and necessity. The notion of a set's containing itself as member is, to many, pathological because sets are (often) conceived as collections of entities. The members must exist independently of the set in order to be collectible.[7] One has analogous intuitions with belief. In the sentence 'you believe this statement', no non-pathological content has been proposed for one's assent or dissent.

All of these derivative notions lead to paradox if their derivative character is ignored and normal principles governing inferences between levels (reflexion principles or principles about self-consciousness) and subject-matter-free transformations (principles of logic or principles characterizing rationality) are respected.

The pathological nature of sentences like (α) or 'you do not believe this proposition' has led many to say that they do not express propositions. From this claim it is sometimes argued that the epistemic paradoxes cannot get off the ground.[8] Similar reactions may be found in the literature on the semantical paradoxes. I do not propose to discuss this reaction in detail here. But I do wish to caution against being caught up in it too blithely.

In the first place, many sentences that lead to paradox are not identifiable as pathological except by reference to empirical facts. And these facts may be unknown to a potential asserter or believer. For example, suppose that Plato thinks that his friend in room 13 is considering the view that the forms are a figment of an overactive imagination. Plato then thinks to himself: I (Plato) do not subscribe to the thought being considered in room 13. Unfortunately, Plato has erred. He himself is in room 13, not his friend. Applying principles like (22) and (23) to this case leads to paradox. Now it is possible simply to deny that Plato could have thought the thought I have described him as thinking.[9] But such a denial is extremely implausible. It seems clearly possible to make statements, hold beliefs, engage in doubts, or experience fears that are expressed by sentences which may lead to semantical or epistemic paradox. Surely a Cretan, not realizing that he is a Cretan in context C, could say or believe that everything said or believed by a Cretan in context C is not true (where all other statements or beliefs in such a context are in fact not true).

In the second place, too facile an appeal to the view that the relevant pathological sentences do not express propositions fails to touch certain paradoxes (Richard's and the paradoxes of grounding) and ignores questions raised by others. Under what conditions, precisely, do sentences fail to express propositions? Why do they fail? How should one equip semantical theory to handle the meaningful, but pathological sentences? These questions and others are lost in the shuffle when propositions are invoked to dismiss the problem.

III

Buridan's solution differs from both the hierarchical approach and the one just deplored. The solution has attracted sporadic interest in modern times – having been developed by A. N. Prior and C. S. Peirce – and is *prima facie* attractive because of its simplicity.[10] Buridan claims that the truth conditions of a sentence (or proposition) are not fully stated by the sentence itself. Those truth conditions also include a statement which is 'virtually implied' (Peirce says 'presupposed') to the effect that the original sentence (or proposition) is true. Both the original sentence and the virtually implied one must be *factually satisfactory* ('it must be as is signified' by both) in order for the original sentence to be true.

The point of this proposal becomes evident when one considers its effect on the liar paradox. Suppose (α_1) is '(α_1) is not true'. Then instead of plugging this sentence alone into scheme (T) and proceeding to a contradiction, one plugs in both (α_1) and its virtual implication on the right side to get

'(α_1) is not true' is true if and only if (α_1) is not true and (α_1) is true.

From this it follows that (α_1) is equivalent to a contradiction and hence is not true. No paradox arises. Thus although (α_1) is meaningful and factually satisfactory – 'things are howsoever (α_1) signifies' –, it is not true.

The application of the approach to the epistemic paradoxes is as follows. In the case of each of the pathological sentences (say, the one expressing a_2, Section II), the supposition that it is true leads to a contradiction. But the supposition that it is not true does not lead to a contradiction. So the sentence is not true. Both the sentence and its negation are counted untrue. The relevant instance of (T) is

(BT) $\quad a_2$ is true if and only if $(-B(P, a_2)$ and a_2 is true).

Since on pain of contradiction a_2 is not true, neither '$-B(P, a_2)$' nor '$B(P, a_2)$' is deducible. Nevertheless, since P (Plato) is most wise, he will know that a_2 is not true and will not believe a_2. Thus a_2 is factually satisfactory, but neither it nor its negation is true.

Since Buridan thought of propositions, or belief contents, as sentence tokens, he had the beginning of an explanation for this situation. When *we* say (truly) that Plato does not believe a_2, we are stating a different proposition from a_2 itself. Both propositions are factually satisfactory; but only ours is true, since a_2 is self-referential in a way that leads to contradiction on the supposition that it is true.

Buridan's approach is unsatisfactory on a number of grounds, of which I will set out the most important. There is a motivational problem. Why should 'virtual implications' be included on the right side of the truth schema in specifying a sentence's truth conditions? Other implications or presuppositions need not be included. If its virtual implication were part of what a sentence asserted or meant, the inclusion would be understandable. But Buridan effectively argues the intuitive implausibility of this viewpoint.[11]

A more significant objection is that the solution does nothing to suggest why sentences like (α), '(α) is true', are pathological in a way intuitively analogous to the paradoxical liar sentence. No insight into the derivative nature of semantical and epistemic notions is provided.

A further drawback is the fact that the truth schema as conceived by Buridan is no longer able to provide the basis for a definition of truth or even an informative recursive characterization. An instance like " '$2 + 2 = 4$' is true if and only if both $2 + 2 = 4$ and '$2 + 2 = 4$' is true" tells one nothing interesting about the notion of truth as applied to the sentence. For to understand what it is for the sentence to be true, one must already understand what it is for the sentence to be true!

The excessive weakness of the assignment of truth conditions through instances of (T) emerges in another way. Suppose (b_1) = '(b_2) is not true' and b_2 = '(b_1) is true'. Thus (b_1) says that its own virtual implication (b_2) is not true. Now using the appropriate Buridanian truth schemas for (b_1) and (b_2), we can prove that (b_2) is not true. But we cannot show that (b_1) is not true – even though we have proved that its virtual implication (or presupposition) is not true. More generally, as can be seen from negating the left and right sides of (BT), *nothing* can be deduced from the assumption or conclusion that a sentence is not true. The Buridan-Peirce-Prior solution avoids contradiction only by making no commitments at all to conditions under which a sentence is *not* true.

Finally, Buridan's solution simply shifts the problems of truth onto the notion of signification (or factual satisfactoriness). From Buridan's remarks, it is clear that the conditions for satisfactory signification are expressed in the schema:

it is as '______' signifies if and only if ______.

But then a contradiction can be derived from the sentence (b_3), 'it is *not* as (b_3) signifies'. One can avoid calling the instances of the argument to contradiction *true* by invoking the original device of virtual implication. But the derivation is unaffected. Thus there remain paradoxes of signification.

The notion of signification cannot be simply jettisoned from the theory. For again consider Plato thinking that he is opposing his friend in room 13. Plato writes 'The sentence written in room 13 is not true'. But Plato is himself in room 13, and has written nothing else. Applying the theory, we judge Plato's sentence to be implicitly a contradiction, and thus untrue. But there remains the intuition that since the sentence means that it is not true – and we have just admitted as much –, the sentence is true after all. Buridan captures this intuition with the notion of signification: the sentence though untrue signifies things as they are.[12] Without the notion of satisfactory signification Buridan cannot account for the intuitive facts. With it, the theory still spawns paradox.

Buridan's appeal to signification has the merit of recognizing the intuitive subtlety of the semantical and epistemic paradoxes. He tries to account for both the intuition that the relevant sentences are pathological and untrue, and the intuition that given what they mean, they are true after all. Buridan's failure to tame the paradoxes by appeal to a distinction between truth and satisfactory signification suggests that the intuitions should be explicated in terms of shifts in the extension of a single notion of truth. A sentence is untrue judged in a certain context of application and true judged in another context. Such a view would appeal to indexical and other pragmatic features of the notion of truth as a means of dispelling the semantical paradoxes. But whatever account one gives of truth, the epistemic paradoxes suggest that belief, doubt and knowledge should be treated similarly.

University of California, Los Angeles

NOTES

* I am indebted for suggestions to Marilyn Adams and Robert L. Martin.

1 Jean Buridan, *Sophisms on Meaning and Truth*, trans. by Theodore Kermit Scott (Meredith Publ. Co. New York, 1966), Chapter VIII, Sophisms 13–15. For discussion of the Hangman Paradox, see David Kaplan and Richard Montague, 'A Paradox Regained', *Notre Dame Journal of Formal Logic* (1960), 79–90.

2 Propositions for Buridan are sentence tokens. The development of the paradoxes will not depend on what one takes propositions to be, although some of the solutions we discuss will rest heavily on a particular choice.

3 Alfred Tarski, 'The Concept of Truth in Formalized Languages', *Logic, Semantics, Metamathematics*, trans. by J. H. Woodger. (At the Clarendon Press, Oxford, 1956.) I discuss Tarski's approach, its most popular rival (the appeal to truth-value gaps), and my own view in 'Semantical Paradox' in preparation. My view includes a hierarchy resembling Tarski's, although weaker restrictions than those stated above are appropriate.

4 Kaplan and Montague, *op. cit.*, esp. pp. 87–88; Richard Montague, 'Syntactical Treatments of Modality with Corollaries on Reflexion Principles and Finite Axiomatizability', *Acta Philosophica Fennica* **16** (1963), 153–167. The paradox (0)–(6) is analogous to the most stripped-down paradox about knowledge presented by Kaplan and Montague. They rely on an analog of the somewhat more artificial proposition: Socrates knows the negation of the proposition on the wall. Since no principles are needed to connect knowledge and belief, the only assumptions required are these: (i) nega $(b) = \overline{-K(S, \text{nega } (b))}$; (ii) $K(S, \overline{-K(S, \text{nega } (b))}) \rightarrow -K(S, \text{nega } (b))$; (iii) $K(S, \overline{\text{(ii)}})$; (iv) $\mathrm{I}(\overline{\text{(ii)}}, \text{nega } (b)) \mathbin{\&} K(S, \overline{\text{(ii)}}) \cdot \rightarrow K(S, \text{nega } (b))$. Jointly these assumptions are contradictory. They are analogous to (0), the left-right direction of (3), (4), and (5) respectively.

5 This remark is somewhat tendentious in that Montague regards his paper as showing that necessity is *unlike* truth. He takes this view because the most straightforward predicate analogs of the iterative principles of modal logic (where necessity is expressed not as a predicate but as a sentential operator) find no place in a hierarchical system: iteration in such a system standardly involves some sort of ramifications (Systems are possible where not all iteration of semantical or modal predicates involves ramification; but some iterations seem to require ramification on any hierarchical approach.) This point, however, does not suffice to establish Montague's view. For no reason is given for thinking that one must find simple, unramified analogs of the iterative principles of modal logic in any adequate account of necessity. To be sure, if one is convinced of the truth of principles like 'If it is necessary that ________, then it is necessary that it is necessary that ________', one must be able to formulate them in one's account of necessity. But to assume that these principles can only be formalized in the unramified way that modal logic formalizes them is tantamount to begging the question – if the question is whether necessity should be treated on an analogy to truth. Montague does not dispute the view that truth should be accounted for hierarchically.

6 In lectures which I first heard in 1974, Alonzo Church discussed certain paradoxes that take the following form. All Bouleus' beliefs except perhaps the following are true. Bouleus' troublesome belief is that at least one of his (Bouleus') beliefs is not true. The resulting paradox can be made to focus on belief in a Russellian language with propositional variables and no truth predicate. The initial assumption is then formalized:

$$(p)\,(\mathrm{Bel}(\mathrm{Bouleus}, p) \mathbin{\&} p \equiv (\mathrm{E}q)(\mathrm{Bel}(\mathrm{Bouleus}, q) \mathbin{\&} -q) \colon \rightarrow p),$$

Where '$\equiv$' represents a strong version of intensional equivalence. This Russellian formalization seems to me misleading. In eschewing the truth predicate, the formalization suggests that the paradox is one specifically concerning belief, whereas the natural-language statement of it gives truth an essential role. In its more orthodox formalizations, as of course Church recognizes, the paradox can be disarmed by simple appeal to Tarski's restrictions on truth. Our paradox above does not submit to such treatment. Cf. Church, 'Comparison of Russell's Resolution of the Semantical Antinomies with that of Tarski', *Journal of Symbolic Logic* **41** (1976), 747–760, Note 25.

[7] Cf. George Boolos, 'The Iterative Conception of Set', *The Journal of Philosophy* **LXVIII** (1971), 215–231; Joseph R. Schoenfeld, *Mathematical Logic* (Reading Mass., Addison Wesley, 1967), Chapter IX. Bertrand Russell was perhaps the first modern writer to fix upon the notion of derivativeness. For obscure informal attempts to express it, see *Principia Mathematica*, second edition (Cambridge, 1927), pp. 39, 54.

[8] Cf. William Kneale, 'Propositions and Truth in Natural Language', *Mind* **81** (1972), 225–243. Tarski himself counts the formal analog of (α) ill-formed, and thus incapable of expressing a proposition. But the hierarchical approach need not, I think, depend on this feature.

[9] The Plato example is adapted from one in A. N. Prior, 'On a Family of Paradoxes', *The Notre Dame Journal of Formal Logic* **2** (1961), 16–32. Prior takes the view here criticized. I intend to discuss this nest of issues further in a future paper. Prior makes several interesting observations about epistemic paradoxes. But all the cases he cites, not excepting the one from which I have derived the Plato example, can be handled by placing restrictions on the notion of truth. This point is somewhat obscured by the fact that Prior treats 'believes' as an operator on sentences: he formalizes his examples in such way that the notion of truth, which explicitly occurs in the natural-language examples, is suppressed in the formalizations. I think there are reasons for not taking the operator approach to belief sentences, but I shall not discuss them here. It is enough to note that the suppression of the notion of truth in the formalization seems unnatural (cf. my remarks regarding the Russellian formalization, Note 6). Part of the interest of our examples is that there is no direct reliance on the notion of truth at all.

[10] Buridan, *op. cit.*, e.g. pp. 92, 194–196; A. N. Prior, 'Some Problems of Self-Reference in John Buridan', *Proceedings of the British Academy* **48** (1962), 281–296; C. S. Peirce, *Collected Papers*, Hartshorne and Weiss (eds.), 5.340, 3.447. Cf. also D. Golstick, 'Against "Categories" ', *Philosophical Studies* **26** (1974), 337–356.

[11] Buridan, *op. cit.*, pp. 194–195. Buridan points out that intuitively not all sentences mean something that involves self-reference. Prior in 'Some Problems', *op. cit.*, champions (what he admits to be) the unintuitive viewpoint as superior to Buridan's because it avoids the paradox of signification discussed below. But Prior's view does not avoid our other objections, and its unintuitiveness prevents it from explicating the intuitions that the notion of satisfactory signification is invoked to explicate. (See below.) It should be noted that Buridan's account requires some technical development in view of the fact that a sentence and its negation can both be untrue.

[12] Neither Peirce nor Prior says anything about this intuition. (Cf. Note 10.) Recent truth-value gap approaches are also virtually unanimous in underrating it. An exception is that of Hans Herzberger, 'Dimensions of Truth', *The Journal of Philosophical Logic* **2** (1973), 535–554. Incidentally, Herzberger received some inspiration from Buridan.