As Noonan agrees, animalism and neo-Lockeanism are incompatible. And the standard neo-Lockean view, as he agrees, is unacceptable. But his new neo-Lockean proposal fares no better. The anti-Lockean animalist's view, I conclude, is the only serious contender.[3]

Oriel College, Oxford

[3] I am very grateful to Penelope Mackie for extensive discussion of, and comments on, the arguments in this paper; also to Paul Snowdon, Derek Parfit, Elena Mitrophanous, Joan Mackie, James Hordern and an anonymous referee of this journal for their comments on an earlier version.

THE TROUBLE WITH HARRISON'S 'THE TROUBLE WITH TARSKI'

By Daniel R. Boisvert

In 'The Trouble With Tarski', *The Philosophical Quarterly*, 48 (1998), pp. 1–22, Jonathan Harrison attacks Tarski's theory of truth and similar theories. According to Harrison, Tarski's theory of truth must be mistaken, since (a) truth cannot be a property of sentences; (b) if truth could be a property of sentences, T-sentences would have to be necessary truths, which they are not; and (c) not only are T-sentences not necessarily true, they can be false. Harrison's attack fails because he fails to understand important elements of a 'Tarski-style' truth-theory.

My first section is a brief review of Tarski's theory of truth and of extensions of his techniques to natural languages. The second is a defence of Tarski-style truth-theories against Harrison's three objections, in three corresponding parts. I confine my remarks to one-fifth of Harrison's article, since the remainder consists mostly of claims which rely on these objections.

I. TARSKI-STYLE TRUTH-THEORIES

Tarski articulated a constraint, Convention T, which must be met by any definition of 'truth' if it is to be materially adequate and formally correct.[1] It can be summarized as follows:

> A definition of a predicate 'is true in *L*' is an adequate definition of a truth-predicate for *L* if and only if one can derive from it all instances of the T-schema '*S* is true in *L* if and only if *p*', where '*S*' is replaced by a structural description of an object-language sentence (sentence of *L*) and '*p*' is replaced by an accurate translation of it into the meta-language (the language in which we are talking about the object-language sentence).[2]

In the T-schema

T. *S* is true in *L* if and only if *p*

the truth-predicate is necessarily relativized to a language if it is applied to sentences, since the same string of symbols can have different meanings in different languages; 'Snow is white' can be true in one language and false or meaningless in another (for example, in a language in which the word 'white' means *black*). Thus '"Snow is white" is true' is akin to 'John is taller'; both are nonsensical unless relativized, to a language in the first case and to another object or class of objects in the second.

Harrison gives (p. 18) what appears to be an argument against the notion of relativization of the truth-predicate. It is quite uncharitable, however, to take Tarski to be using 'statement' as synonymous with 'proposition',[3] and to commit Tarski to an elementary use/mention confusion, something to which he was not prone.

Harrison (pp. 1–2, 21–2) particularly attacks Tarski's work in application to natural languages. (Hereafter I use 'T-theory' to denote Tarski-style truth-theories for both formalized and natural languages, i.e., truth-theories for natural languages meeting a suitably modified Convention T, and 'T-theorist' for a supporter of T-theories.) To define a truth-predicate in Tarski's style for a natural language, the object-language sentence described by what replaces '*S*' in the T-schema must also be relativized to a speaker and time so as to accommodate context-sensitive elements

[1] For characterizations of 'materially adequate' and 'formally correct', see A. Tarski, 'The Concept of Truth in Formalized Languages' (1933, hereafter TFL), repr. in his *Logic, Semantics, Metamathematics* (Oxford UP, 1956), pp. 152–278, at pp. 155, 165. See also Tarski's 'The Semantic Conception of Truth and the Foundations of Semantics' (1944, hereafter SCT), repr. in A.P. Martinich (ed.), *The Philosophy of Language*, 3rd edn (Oxford UP, 1996), pp. 61–84, at pp. 61, 62–4.

[2] For the more technical definition of 'Convention T' see TFL pp. 187–8. For a more thorough discussion of Convention T, as well as of the object-language/meta-language distinction, to be discussed below, see J. Alberto Coffa, *The Semantic Tradition from Kant to Carnap: to the Vienna Station* (Cambridge UP, 1991), pp. 293–300.

[3] See TFL p. 156 fn. 1, where Tarski states explicitly that he takes 'statement' to be synonymous with 'sentence', and SCT p. 62, where Tarski uses 'sentence' when discussing the point of the passage mentioned by Harrison.

in natural languages. Hence the T-schema must be modified in the following style if one wishes to define 'truth' for a natural language[4] (letting '*u*', for 'utterer', range over speakers, and '*t*' over time intervals):

T′. $(\forall u)(\forall t)$(*S*, taken as if actually spoken by *u* at *t*, is true in *L* if and only if *p*).

Let (TN) abbreviate (T′):

TN. *S* is $\text{true}_{[u,t]}$ in *L* if and only if *p*.

Thus Convention T adapted to a T-theory for a natural language requires that the truth-predicate be relativized to a language, and an object-language sentence to a speaker and time.

The T-schema involves an object-language and a meta-language (SCT p. 67), a distinction required to avoid semantic paradoxes (SCT pp. 65–6, TFL pp. 157–8). Hence all instances of the T-schema are written in the meta-language, which must

> contain an adequate translation of the object-language sentence under consideration
> be rich enough to provide a structural description for any sentence in the object-language
> contain logical terms
> be essentially richer than the object-language, i.e., it must contain at least the means to construct a structural description for any object-language sentence, and the truth-predicate for the object-language (which is not contained in the object-language).

What replaces '*S*' in (T) is a *description* of an object-language sentence in terms of its primitive components. Thus one must not suppose, as Harrison (p. 16) does, that what composes (in part) an instance of '*S*' must also compose an instance of '*p*', even if the meta-language embeds a context-insensitive object-language. Hereafter I use Quinean corner-quotes to form a description of an expression in terms of its components when formulating instances of the T-schema.

II. THE TROUBLE WITH HARRISON

1. *The property objection: truth is not a property of sentences*

Harrison's first objection (pp. 2, 3–4), which I shall call 'the property objection', can be summarized as follows: (P1) T-theories must treat sentences as truth-bearers; but (P2) truth cannot be a property of sentences; thus (P3) T-theories must be misguided. The property objection is unsound because (P1) is false and (P2) is unjustified.

Harrison assumes that T-theorists commit themselves to sentences as truth-bearers, and he cites a passage (TFL p. 157) in which Tarski supposedly so commits

[4] Cf. D. Davidson, 'Truth and Meaning', in his *Inquiries Into Truth and Interpretation* (Oxford UP, 1984), pp. 17–36, at p. 34; G. Evans, 'Does Tense Logic Rest on a Mistake?', in his *Collected Papers* (Oxford: Clarendon Press, 1985), pp. 343–63, at pp. 359–60.

himself. But a T-theory can be formulated in terms of *whatever* truth-bearers might be, whether sentences, beliefs, assertions or propositions. For example, one could employ a T-theory to define 'truth' for beliefs by slightly modifying Convention T so that any materially adequate definition of the truth-predicate 'is true in *L*' entails all instances of

TB. $(\forall u)(\forall t)$(The belief whose content is expressed by *S* taken as if actually spoken by *u* at *t* in *L* is true if and only if *p*)

where '*S*' is replaced by a structural description of an object-language sentence and '*p*' by an adequate meta-language translation of it. Similarly, one could employ a T-theory to define 'truth' for assertions or propositions (indeed, Harrison himself attempts at p. 16 to modify Convention T thus). So T-theorists need not commit themselves to sentences as truth-bearers. Indeed, Tarski did *not* commit himself to this: he states explicitly (SCT p. 62) that he applies the truth-predicate to sentences as a matter of convenience, and he is clear that it may be applied to other entities, including propositions. (P1) is certainly false, and therefore the property objection is unsound for this reason alone.

But Harrison also fails (p. 3) to justify (P2):

> It follows from this fact [that the sentence 'Snow is white' might be 'false' although snow was white] that truth is not a 'property' of the sentence 'Snow is white', for (if truth and falsehood could be predicated of the sentence at all) the sentence could be false, although snow was white. But though the sentence 'Snow is white' might be 'false', though snow was white, it cannot be false *that* snow is white, if snow is white. It may well be a necessary truth – and will be a necessary truth, three-valued logics apart – that that snow is white is true, if snow is white.

Here the object-language sentence 'Snow is white' is not relativized to a speaker and a time, nor is the truth-predicate relativized to a language. Hence this argument fails even to engage T-theorists. But, even if so relativized, the argument fails, for it relies on the false assumption that T-sentences can be false ('the sentence "Snow is white" could be false, although snow was white'). 'Snow is white' *could* be false relative to a language and/or a time although snow was white (to take Harrison's example, in a language in which 'white' meant *black*), but this does not support the assumption that T-sentences can be false. For if in a language *L* 'white' meant *black*, the sentence (in English)

1. ⌜Snow is white⌝ is $\text{true}_{[u,t]}$ in *L* if and only if snow is white

would *not* be a T-sentence, since its right-hand side would not be an adequate meta-language translation of the object-language sentence 'Snow is white'. As we shall see, no T-sentence is false; hence Harrison also fails to justify (P2). (Harrison also attempts at pp. 3–4 to 'reinforce' (P2), but this passage is fraught with difficulties: it (a) begs the question; (b) unjustifiably shifts the burden of proof; and (c) relies on unargued but controversial ontological commitments.) Consequently the property objection is unsound and poses no threat to T-theorists.

2. *The necessity objection: T-sentences are not necessary truths*

Harrison's second objection (pp. 1–2), which I shall call 'the necessity objection', can be summarized as follows: (N1) according to T-theorists, any materially adequate definition of 'true' must entail T-sentences; (N2) these are necessary truths; but (N3) T-sentences are not necessary truths; thus (N4) T-theories are misguided. This objection fails, since it relies either on a classic modal fallacy or on an equivocation.

Here is Harrison's brief 'argument' (p. 1) for thinking that T-theorists must accept (N2):

> It ought to be held on pain of inconsistency by all [T-theorists] ... that sentences like '"Snow is white" is true if and only if snow is white' are necessary truths. For it *is* a necessary truth ... that whatever is made true by snow's being white, whether it be a sentence or a proposition, must be true if snow is white. (Necessarily: *whatever is made true by snow's being white is true if snow is white* – because each entails the other; not: *if snow is white, whatever it makes true is necessarily true.*)

Again, since neither the object-language sentence 'Snow is white' nor the truth-predicate are appropriately relativized, this argument fails to engage the T-theorist. Moreover, the argument seems to rest on the classic modal fallacy of confusing the necessity of the consequent with the necessity of the consequence (in the numbering below, from (2) and (3) to (6)). Supplying the appropriate relativizations and taking English as the language of interest permits a reconstruction of the argument which does not rest on this modal fallacy and which, if sound, supports (N2):[5]

2. Necessarily, for any sentence x, if x is made $\text{true}_{[u,t]}$ in English by snow's being white, then x is $\text{true}_{[u,t]}$ in English if and only if snow is white [assumption]
3. ⌜Snow is white⌝ is made $\text{true}_{[u,t]}$ in English by snow's being white [assumption]
4. For any sentence x, if x is made $\text{true}_{[u,t]}$ in English by snow's being white, then necessarily, x is made $\text{true}_{[u,t]}$ in English by snow's being white [tacit assumption]
5. Necessarily, ⌜Snow is white⌝ is made $\text{true}_{[u,t]}$ in English by snow's being white [from (3) and (4)]
6. Necessarily, ⌜Snow is white⌝ is $\text{true}_{[u,t]}$ in English if and only if snow is white [from (2) and (5)].

This argument is sound provided that 'English' fixes the same semantic properties for its sentences in all possible worlds. To illustrate, if 'English' is used to fix rigidly the semantic properties for its sentences, then no word in the **English** language (using bold face to indicate use of 'English' to designate the language's semantic properties rigidly), so understood, could ever have a different meaning; thus 'white' could not mean *black* in **English**. Consequently the T-sentence

7. ⌜Snow is white⌝ is $\text{true}_{[u,t]}$ in **English** if and only if snow is white

would be a necessary truth, for there could be no circumstance in which 'Snow is white' could be false in **English** if snow was white, and no circumstance in which

[5] I thank Greg Ray for suggesting the form of this reconstruction.

'Snow is white' could be true in **English** if snow was not white. (I suppress here relativization to speaker and time.) Thus, if 'English' fixes rigidly the semantic properties of the language, then T-sentences are necessary truths.

However, if 'English' is not used to fix rigidly the semantic properties of its sentences, then it is possible for words to have meant, and to come to mean, something other than what they now mean in relation to what 'English' denotes. For example, 'English' is often used to designate the language (in the philosopher's or logician's sense) spoken at any time by members of a certain linguistic community, whose linguistic practices can change. Suppose 'English' is used to mean 'The language spoken by most Anglo-Saxons' (when I use 'English' in this sense, I do not use bold face). Thus most Anglo-Saxons today use 'white' to mean *white*. But if they came to use 'white' to mean *black*, say in 2025, then the English sentence 'Snow is white' would be true in 1998 English even though it would be false in 2025 English. Consequently (7), modified so that 'English' is used non-rigidly, would not be a necessary truth, for there could be circumstances in which 'Snow is white' is false in English even though snow was white. I stress two points. First, such a circumstance is the basis of the 'counter-examples' that Harrison uses to justify (N3). Second, even though such a circumstance would make (7) false, it would not make a *T-sentence* false, since in such a circumstance (7) would not be a T-sentence (because the right-hand side of the biconditional would not be an adequate meta-language translation of the English sentence 'Snow is white'). The T-sentence (in 1998 English) for 'snow is white' in 2025 English would be

8. ⌜Snow is white⌝ is $\text{true}_{[u,t]}$ in English in 2025 if and only if snow is black.

This (given our hypothetical) would be true, though not necessarily so (for the reason just given). Thus there are no *false* T-sentences; all are true, and necessarily true if the term denoting the language to which the truth-predicate is relativized is treated as fixing rigidly the semantic properties of its sentences.

What are the implications for the necessity objection? Harrison claims that T-theorists must hold that T-sentences express necessary truths, but in fact T-sentences do not express necessary truths, so that T-theorists are committed to something that is false. But for this objection even to look sound – taking now into account the necessary relativization of the truth-predicate to a language – we would have to equivocate in the use of the term used to denote the object-language in the T-sentence. Let λ be the term denoting the object-language. For the T-theorist to be committed to (N2), he must take λ to fix rigidly the semantic properties of its sentences. However, on the assumption that λ does so, (N3) is manifestly false. For (N3) to be true, we must suppose that λ does not fix rigidly the semantic properties of its sentences. But if we suppose this, then the T-theorist need not be committed to T-sentences' expressing necessary truths. Thus whether we take λ to fix rigidly or non-rigidly the semantic properties of its sentences, the objection fails. If it is taken to fix the semantic properties rigidly, the first assumption, that the T-theorist is committed to (N2), is true, but (N3) is false. If it is taken to fix these properties non-rigidly, then (N3) is true, but the assumption that the T-theorist is committed to (N2) is false. It is only if we take it to fix rigidly the semantic properties in considering

(N2) but not to fix them rigidly in considering (N3) that the argument looks sound; but that is an equivocation. Consequently the necessity objection poses no threat to T-theorists.

3. *The falsity objection: T-sentences can be false*

Harrison's third objection (p. 2), which I shall call 'the falsity objection', can be summarized as follows: (F1) according to T-theorists, any materially adequate definition of 'true' must entail T-sentences; (F2) these are at least true; but (F3) T-sentences can be false; thus (F4) T-theories are misguided. This objection fails because (F3) is false.

T-theorists are committed to (F2), for, as we have seen, a sentence of the form (T) is not a T-sentence unless the sentence on the right-hand side translates the object-language sentence described on the left-hand side, and this guarantees its truth.[6] But (F3) is false. Harrison tries to justify (F3) in two ways, first by offering counter-examples to (F2). He again invokes the possible circumstance in which the English sentence 'Snow is white' is false in English because the word 'white' means *black* even though snow is white. In such a case, (7), repeated here (and understood in 1998 English, but thinking of 'English' in (7) as relative to, say, 2025)

7. ⌜Snow is white⌝ is $\text{true}_{[u,t]}$ in English if and only if snow is white

would be false, since the left-hand side of (7) is false and the right-hand side true. In that case, (7) would indeed be false; but then (7) would not be a T-sentence because the right-hand side of the biconditional is not an adequate meta-language translation of 'Snow is white'. Thus Harrison's first justification is inadequate because he fails to provide any counter-examples to (F2).

Harrison's second justification for (F3) is this (p. 2):

> the sentence ['"Snow is white" is "true" if and only if snow is white'] ... is not even true. For ... snow has been white for much longer than there has been an English sentence 'Snow is white', and will remain white after the English language has ceased to exist – say, because in the future we will all become extinct or speak Chinese.

Again, since neither the object-language sentence nor the truth-predicate are appropriately relativized, this argument fails to engage T-theorists at all. But even supplying the relativizations, this argument fails because it mistakenly assumes that languages exist only when there are populations that speak them. A language *L* might be thought of as a function mapping meanings to those strings of symbols which are the sentences of *L*.[7] This function is a set of ordered *pairs* of meanings and strings of symbols and does not require any human involvement. Hence languages so construed are independent of human involvement, and therefore it is false that

[6] For an explanation of a slight sophistication required if the object-language sentence contains context-sensitive elements, see E. LePore and K. Ludwig, 'Outline of a Truth Conditional Semantics for Tense', in Q. Smith (ed.), *Tense, Time and Reference* (Oxford UP, forthcoming).

[7] Cf. D. Lewis, 'Languages and Language' (1975), repr. in Martinich (ed.), *The Philosophy of Language*, pp. 538–57, at p. 542.

languages exist only when there are populations that speak them. Harrison again fails to justify (F3), and consequently the falsity objection poses no threat to T-theorists.

III. CONCLUSION

Tarski's project was to provide a materially adequate and formally correct definition of the notion of truth. In order to ensure the satisfaction of this first condition, he articulated a constraint, Convention T: a materially adequate definition of 'truth' entails all T-sentences, which are instances of the T-schema '*S* is true in *L* if and only if *p*', where '*S*' is replaced by a structural description of an object-language sentence and '*p*' is replaced by an accurate meta-language translation of it. A T-theory is amenable to whatever bears the truth-predicate, be it beliefs, assertions, propositions or sentences (if relativized to a language). For this reason, the property objection fails. T-sentences are always true, and necessarily true if λ fixes rigidly the semantic properties of the sentences of that language. Thus the necessity objection and the falsity objection also fail. Finally, the falsity objection also fails because Harrison mistakenly assumes that languages exist only when there are populations that speak them. These objections pose no threat to T-theorists, and thus Harrison has not made any trouble at all for Tarski.[8]

University of Florida

[8] I thank Kirk Ludwig and Greg Ray for helpful suggestions on an earlier version of this paper.