more boldly, simply being in a certain sort of attitudinal state, that matters. Thus for both theoretical and moral reasons, it is better, given how the sensation sorts crosscut the attitudinal criteria, to take these latter, not the former, as definitional of 'pain'. Pain is an attitude, not a sensation.

NORTON NELKIN

University of New Orleans

BRAINS IN A VAT*

IN chapter 1 of *Reason, Truth, and History,*[1] Hilary Putnam argues from some plausible assumptions about the nature of reference to the conclusion that it is not possible that all sentient creatures are brains in a vat. If this argument is successful, it seemingly refutes an updated form of Cartesian skepticism concerning knowledge of physical objects. In this paper, I will state what I take to be the most promising interpretation of Putnam's argument. My reconstructed argument differs from an argument strongly suggested by Putnam's text. I will show that the latter argument obviously does not work. The more promising argument which I reconstruct on behalf of Putnam raises some interesting questions about the relation between the contents of one's beliefs and one's environment and about how this relation affects the evaluation of anti-skeptical arguments. I conclude that my reconstructed argument ultimately fails as a response to Cartesian skepticism: the argument engenders a skepticism about knowledge of meaning, or propositional content, which undercuts its anti-skeptical force.[2]

* I would like to thank the members of a seminar on epistemology at Yale University in fall, 1983, for helpful discussions of these issues. I have also benefited from conversations with John Fischer and Jonathan Wilwerding. I am especially indebted to Phillip Bricker for quite extensive criticisms and suggestions which greatly changed and improved this paper.

[1] New York: Cambridge University Press, 1981; parenthetical page references will be to this book, unless otherwise noted.

[2] The argument of chapter 1 should be sharply distinguished from the "model-theoretic" argument against metaphysical realism which Putnam develops in chapters 2 and 3 of his book [see also his "Realism and Reason" in *Meaning and the Moral Sciences* (Boston: Routledge & Kegan Paul, 1978) and his "Models and Reality," *Journal of Symbolic Logic,* XLV, 4 (September 1980): 464–482)]. His argument against metaphysical realism, if successful, would show, in a quite different way from chapter 1's argument, that the brains-in-a-vat "possibility" is incoherent. The argument of chapter 1 indeed depends upon causal-theoretic assumptions

0022-362X/86/8303/0148$02.00

I

I will begin by stating a Cartesian skeptical argument about brains in a vat. Let us say that if Q is a logically possible proposition that is incompatible with P and P is a logically possible proposition, then Q is a *counterpossibility* to P. Let us also state a *counterpossibility principle:*

(CP) If I know that P and that Q is a counterpossibility to P, then I know that Q is not the case.[3]

The argument proceeds as follows.

(A) *That I am a brain in a vat inhabiting a world in which the only objects are brains in a vat and laboratories containing computers programmed to stimulate the brains* is a logically possible proposition.
(B) If I am a brain in a vat of the Putnamian sort just specified (hereafter a *BIV*), then I am not, for example, now sitting on a chair.
(C) The proposition that I am a BIV is a counterpossibility to the proposition that I am now sitting on a chair. [(A), (B)]
(D) If I know that I am now sitting on a chair and that the proposition that I am a BIV is a counterpossibility to the proposition that I am now sitting on a chair, then I know that I am not a BIV. [(CP)]
(E) I know that (C).
(F) I do not know that I am not a BIV.
(G) I do not know that I am now sitting on a chair. [(D), (E), (F)]

The same argument can be stated with respect to every proposition about physical objects which I claim to know, except the propositions that there are objects, that there are computers, that there are brains, that there are vats, and the like (propositions that would be true even if I were a BIV). Now if Putnam can show that it is *not* possible that all sentient creatures are BIVs, then he can block the foregoing argument by refuting premise (A). This is indeed the kind of anti-skeptical strategy which is suggested by many of Putnam's

about reference which Putnam explicitly rejects in chapters 2 and 3. Putnam has indicated (in conversation) that it was in fact his intention to construct an argument in chapter 1 quite different from the model-theoretic argument of the later chapters. For a criticism of that argument, see my "Putnam's Model-theoretic Argument against Metaphysical Realism," *Analysis,* XLIV.3, 203 (June 1984): 134–140.

[3] (CP) is not importantly different from the principle that knowledge is closed under known logical implication:

If I know that P and that P logically implies Q, then I know that Q.

One might challenge the skeptical argument by challenging this sort of principle, but this is not Putnam's strategy.

remarks, but later in this section I will show that it is not available to him.

Putnam's argument to show that (A) is false, that is, that it is not possible that I am a BIV, depends upon an analysis of the truth conditions for the sentence 'I am a BIV' as uttered (or thought) by a BIV.[4] It is natural to suppose that the sentence would be *true* as uttered by a BIV and that, correlatively, the sentence 'I am sitting on a chair' would be *false* as uttered by a BIV. The BIV's utterance of 'I am a BIV' would presumably be held to be true, though, by virtue of the supposed fact that a BIV's token of 'brain' would refer to brains and his token of 'vat' would refer to vats. However, Putnam argues that a proper understanding of the causal requirements for reference will show that the brain's token of 'BIV' does not refer to BIVs.

Suppose that there are no trees on Mars and that a Martian forms a mental image exactly resembling one of *my* tree images as a result of his perceiving a blob of paint that accidentally resembles a tree. Putnam's intuition is that the Martian's image is not a representation of a tree. Now if I were a BIV, then (a) my mental life would be qualitatively identical to my actual mental life ("from the inside"), (b) my mental life would be caused by a computer's electrical stimulation of my brain, (c) the same would be true of every other sentient creature, (d) the situation described in (a)–(c) would have arisen completely randomly, and (e) there would be no trees, since there would be no objects other than brains, a vat, and the computers that stimulate the brains. If I were a BIV, then, my mental image "of a tree" would no more be a representation of a tree than would the Martian's mental image. Neither of us would have the sort of causal contact with trees which is required for our images to refer to trees. The same reasoning applies to any tokens of 'tree' which might come to be uttered (or thought) by the Martian and by my BIV counterpart. (We assume that the Martian's token is just as randomly caused as is his "treelike" image—we assume that the token is not ultimately caused by, say, a conversation with a visiting earthling.)

What does the BIV's token of 'tree' refer to, if not to trees? Putnam offers three possibilities: (i) to "trees-in-the-image" (I take it that by 'the image' Putnam means *the succession of sense impressions had by the BIV*), (ii) to the electrical impulses that stimulate the brain and thereby cause it to have sense impressions just like those I have when I see a tree, and (iii) to the program features that are causally respon-

[4] In speaking about BIVs, I will use 'utter' to mean, in effect, 'seem to utter', since a BIV cannot speak or write, but only seems to himself to be speaking or writing. Alternatively, one could take 'utter' to mean 'think a sentence token'.

sible for the stimuli described in (ii). On the *natural* assignment of references which one would make in evaluating the truth value of a BIV's utterance of 'Here is a tree', one would hold that the brain's token of 'tree' refers to trees and, hence, that his sentence token is false (expresses a falsehood),[5] since he is not near a tree. On each of Putnam's proposed reference assignments, though, the brain's sentence token comes out *true* (provided that the brain is indeed being stimulated so as to have sense impressions just like those I have when I see a tree and that the stimulation is caused by a computer's program features). On account (i), e.g., the brain's utterance of 'Here is a tree' is true iff the brain is having sense impressions as of being near a tree.[6]

On account (i), a BIV's token of 'BIV' refers to *BIVs-in-the-image,* and a BIV's utterance of 'I am a BIV' would be true iff he were a BIV-in-the-image. As I understand it, those truth conditions are equivalent to these: the BIV's utterance would be true iff he had sense impressions as of being a BIV. But by Putnam's hypothesis, a BIV never has such sense impressions. A BIV has only sense impressions as of being a normal, embodied human being moving through a richly varied world of physical objects. Thus, on account (i), a BIV's utterance of 'I am a BIV' would never to true, contrary to the deliverance of what I called the "natural" account of the utterance's truth conditions.[7]

Before considering the question of what actually follows from Putnam's claims about reference, we must note that these claims do

[5] It makes no difference to this paper whether sentence tokens or propositions (or something else) are the truth-bearers. I will often speak of *utterances* as being true. One could regard this as shorthand for whichever account is deemed superior.

[6] More exactly, the truth conditions for a BIV's utterance of 'Here is a tree' are the would-be phenomenalist truth conditions for English utterances of the sentence. That is, a sophisticated phenomenalist would hold that, e.g., one need not have sense impressions as of a green tree in order for one's sentence 'Here is a green tree' to be true. This is because a sophisticated phenomenalist would want to allow for the drawing of an is/seems distinction within the phenomenalistic language, allowing one to say truly, "A green tree is before me, though I have sense impressions as of a red tree (though there seems to be a red tree before me)." The truth conditions for a BIV's utterance of 'I am a BIV' [on account (i)] accordingly concern more complex facts about sense impressions than those discussed in the text. However, please allow me the more simplistic formulation in the text in order to avoid excessive complication. I am indebted here to David Braun and Michael Thompson.

[7] On account (i), a person whom we would normally consider to be a victim of deception by a nonphysical Cartesian evil genius would say something true when he uttered 'Here is a tree', since 'tree' would refer to sense impressions as of trees. On the analogue to accounts (ii) and (iii), the victim's term would presumably refer to the states of the evil genius which are causally responsible for the aforementioned sense impressions, and the victim would again say something true. (So he would not really be a *victim of deception.*)

not obviously hold for what one might call a *standard* brain-in-a-vat situation. The claims do not obviously hold for a situation in which I am a brain in a vat and am stimulated by evil neuroscientists who, e.g., stimulate my brain in exactly the way their brains are stimulated when they see a beech tree outside their laboratory, thereby causing me to have sense impressions just like their *veridical* sense impressions as of beeches. In such a standard brain-in-a-vat situation, a reasonable causal theorist of reference would surely hold that the brain's token of 'tree' is causally connected with trees by virtue of the token's causal connection with the evil neuroscientist and *his* tokens of 'tree'. Putnam's claims about reference, then, hold at best for worlds in which there is nothing other than brains in a vat and their automatic tenders (innocent human bystanders causally unconnected with the computers would not affect Putnam's claims). So even if it follows from Putnam's remarks that it is not possible that I in a vat of the latter sort), it does not follow that it is not possible that I am a brain in a vat of the standard sort. Hence Putnam's remarks have no force against a Cartesian skeptical argument built upon the (supposed) counterpossibility that I am a brain in a vat of the standard sort.

Do Putnam's remarks about reference even show that "it cannot possibly be true" that I am a BIV (a *Putnamian* brain in a vat)? Here is Putnam's own summary of his reasoning: "In short, if we are brains in a vat, then 'We are brains in a vat' is false. So it is (necessarily) false" (15). This argument (which depends upon the claims about reference) does not show that the proposition that I am a BIV is not a logically possible proposition. It does not show that this proposition is necessarily false. The main point of Putnam's remarks about reference is that the sentence 'I am a BIV' as uttered by a BIV expresses a *different proposition* from the proposition it expresses as uttered by a non-BIV. In the latter case, the sentence expresses a proposition about BIVs, whereas in the former case, it expresses a proposition about sense impressions [on account (i)]. That is, the truth conditions for utterances of the sentence shift from the one case to the other (since the references of the parts of the sentence shift), and this means that the sentence expresses different propositions in the two cases. So we cannot reason from Putnam's claims about reference to the conclusion that a single proposition—that I am a BIV—is not a logically possible proposition (is necessarily false).

Putnam at times states his conclusion in a less misleading way. He says that the proposition that I am a BIV is, like the proposition that I do not exist, a *self-refuting supposition,* such that the entertaining or enunciating of the supposition entails its falsity (8/9). But this is

not quite right either as an account of what Putnam has argued on the basis of his claims about reference. Being self-refuting, in the sense Putnam apparently has in mind, is a property not of suppositions or propositions, but rather of sentences. It is the following property: if a self-refuting sentence *S* of a given language is uttered or thought, then *S* expresses some false proposition or other. Some self-refuting sentences (such as 'I do not exist') express different propositions when different people utter them, so that in such cases, there is no *single* supposition or proposition that is false simply by virtue of being entertained. Now the proposition expressed by a particular utterance of a self-refuting sentence need not be a necessarily false (logically impossible) proposition.[8] The proposition expressed by a particular utterance of the English sentence 'I do not exist' is obviously not necessarily false. When I utter the sentence, I express a proposition that is true at some possible worlds (those in which I never come into existence).[9] Similarly, for all Putnam has shown, the proposition that I am a BIV is not necessarily false either. It is rather that, if I were a BIV, the proposition that my sentence 'I am a BIV' would express would be false, and, if I were not a BIV, the *different* proposition that my sentence 'I am a BIV' would express in that case would be false as well. It is not even quite right to say that, given all this, the sentence 'I am a BIV' is self-refuting in the way in which 'I do not exist' is. This is because when a non-BIV English speaker utters the sentence, it is apparently part of a different language from that of a BIV, since a BIV speaks vat-English, a language with a great number of semantical properties different from those of English. So there is in this case no sentence *of a given language* which expresses a false proposition when uttered by a non-BIV and expresses a *different* false proposition when uttered by a BIV.[10]

[8] A sentence that expresses a necessarily false proposition will be a self-refuting sentence, on the present account.

[9] On certain views of the semantics of indexical expressions (e.g., David Kaplan's *Demonstratives*, draft #2, mimeograph, UCLA Department of Philosophy), the whole story will not be told if we say that 'I do not exist' expresses different propositions when uttered by different people. This is because (on Kaplan's view, for example) the sentence has a uniform *character* across different speakers, though the *content* shifts. The content, though, is true at some worlds and false at others.

[10] One might hold that a BIV speaks the same language as an English-speaking non-BIV, though many of the BIV's terms have different references from the non-BIV's (just as my indexicals have different references from yours, even though we both speak English). Nothing in this paper hinges upon whether we call the BIV's language English. All that is crucial is that many of the BIV's terms have different references from the non-BIV's, and that one does not know which set of references one's own terms have (the BIV set or the non-BIV set).

II

If Putnam cannot show that the proposition that I am a BIV is necessarily false, then he cannot refute the Cartesian skeptical argument considered above by way of refuting premise (A) in the envisaged manner. So the suggested anti-skeptical strategy seems to fail. If he is right, however, in holding that the sentence 'I am a BIV' has a property like that of being self-refuting (that it has some such property follows from the claims about reference), then Putnam apparently does have available another way of refuting the skeptical argument. Consider the following reasoning, in which I use account (i) of vat-English truth conditions.

(1) Either I am a BIV (speaking vat-English) or I am a non-BIV (speaking English).
(2) If I am a BIV (speaking vat-English), then my utterances of 'I am a BIV' are true iff I have sense impressions as of being a BIV.
(3) If I am a BIV (speaking vat-English), then I do not have sense impressions as of being a BIV.
(4) If I am a BIV (speaking vat-English), then my utterances of 'I am a BIV' are false. [(2), (3)]
(5) If I am a non-BIV (speaking English), then my utterances of 'I am a BIV' are true iff I am a BIV.
(6) If I am a non-BIV (speaking English), then my utterances of 'I am a BIV' are false. [(5)]
(7) My utterances of 'I am a BIV' are false. [(1), (4), (6)]

The intuition behind the argument is that no matter whether or not I am a BIV—no matter whether I am speaking vat-English or English—when I utter 'I am a BIV', I say something false. So, by the argument, I know that whatever proposition is expressed by my utterances of 'I am a BIV' is a false proposition. How does this conclusion enable me to refute the Cartesian skeptical argument? Presumably the proposition expressed by my utterances of 'I am a BIV' is the proposition that I am a BIV. So if I know on the basis of the above argument that the proposition expressed by my utterances of 'I am a BIV' is false, then I know that it is false that I am a BIV. Hence, I have refuted premise (F) of the skeptical argument

(F) I do not know that I am not a BIV.

On this strategy, then, the anti-skeptic does not try to refute the skeptical argument by refuting premise (A), as in the strategy suggested by some of Putnam's remarks and rejected above.

Before proceeding to evaluate the above anti-skeptical argument, I will pause in order to point out the affinities between the reconstructed Putnamian position and an interesting position in the phil-

osophy of mind. Suppose that one holds that in Putnam's famous Twin Earth example, the sentence 'Water is wet' expresses different propositions when my twin and I utter it, simply in virtue of the fact that the liquid on Twin Earth superficially indistinguishable from earthly water has the chemical structure XYZ rather than H_2O. Anyone who holds this view has the makings of an anti-skeptical argument like that above at his disposal. Tyler Burge, for example, in "Other Bodies,"[11] maintains that "the contents of Adam and . . . [Adam's twin's] beliefs and thoughts differ while every feature of their non-intentionally and individualistically described physical, behavioural, dispositional, and phenomenal histories remains the same." On Burge's "anti-individualistic" view, it would, for example, be incorrect for me to attribute to my twin the belief that water is wet. According to Burge, my twin's term 'water' does not mean the same thing as my term 'water' despite our exact similarity from the skin inward, given the difference in our physical environments. It would therefore be incorrect to use my term 'water' in oblique occurrence in a 'that'-clause which ascribes content to the belief that my twin expresses by using 'Water is wet'. It would be best to coin a word like 'twater' to use in such (de dicto) belief ascription.

Burge does not in fact use these claims about how one's physical environment can affect the content of one's thoughts and beliefs in the same anti-skeptical manner as that embodied in the foregoing argument. He does, however, think it would be incredible to claim that my twin would hold beliefs involving the notion of water in a solipsistic world in which no water exists, no twater exists, and no community of speakers exists to spin even deluded stories using the term 'water'. How would my twin have acquired the concept of water, as opposed to, say, the concept of twater in such a world, Burge wonders? Why would the beliefs expressed by his uses of 'water' in such a world be false when evaluated with respect to that world and yet (a) *true* when evaluated with respect to a world containing water, but (b) *false* when evaluated with respect to a world containing twater? Wouldn't it be just as reasonable to hold that the beliefs would be *false* when evaluated with respect to a world containing water, but *true* when evaluated with respect to a world containing twater?[12] Burge's point is apparently that there is no clearly correct

[11] In Andrew Woodfield, ed., *Thought and Object* (New York: Oxford, 1982); see also Burge's "Individualism and the Mental", in P. S. French, T. E. Uehling, and H. K. Wettstein, eds., *Studies in Metaphysics, Midwest Studies in Philosophy,* vol. IV (Minneapolis: University of Minnesota Press, 1979).

[12] See "Other Bodies," pp. 114–118, for Burge's remarks on skepticism. See also fn 18, p. 120. Burge has indicated (in conversation) that he regarded a straightfor-

choice to be made in the ascription of content to a solipsistic subject's "beliefs" expressed using 'water' and that there is hence something incoherent in the supposition that it is possible that I am a solipsistic subject holding the false belief that water is wet (and the false belief that I am sitting on a chair, etc.). This suggestion of an anti-skeptical strategy is echoed by Putnam's remark (15) that a BIV might mean nothing at all when he utters 'I am a BIV'. That is, the suggestion seems to be that there is no clearly correct choice to be made in the ascription of content to a BIV's "beliefs." This sort of anti-skeptical move is importantly different from that embodied in the reconstructed Putnamian argument, according to which there *is* a correct choice: the correct ascription of content to a BIV's beliefs is such as to make them *true*. The evaluation of the variant Burge-style strategy will have to wait upon the evaluation of the reconstructed Putnamian argument.[13]

III

Let us from now on refer to the set of sentences (1)–(7), which express our anti-skeptical argument, as *E*. To begin to see how peculiar E is, consider the following apparent problem with premise (2). It seems that (2) does not correctly state the truth conditions that my utterances of 'I am a BIV' have on condition that I am a BIV. We are assuming that 'true' (and 'if', 'and', etc.) have the same semantical properties in their occurrences in vat-English that they have in their occurrences in English. Hence, even for a BIV, the following sentence would be true as uttered by him in vat-English:

(T) My utterances of 'I am a BIV' are true iff I am a BIV.

(T) is true so long as the metalanguage used in stating (T) is the same as (or contains) the language of the mentioned sentence. And this relation does hold for the relevant meta- and object languages: if (T) is uttered by a BIV, then vat-English will be both the meta- and the object language, and if (T) is uttered by a non-BIV, English will be both the meta- and the object language. It is not as if a speaker might

---

ward anti-individualist argument against skepticism as no more than a tantalizing possibility when he was writing "Other Bodies" and is now dubious about the success of such reasoning.

[13] See Jerry Fodor's "Narrow Content and Meaning Holism" (forthcoming) for a conception of belief content according to which my twin and I express the same content by our utterances of 'Water is wet'. On this view, one cannot develop a Burge-style anti-skeptical strategy. See also Fodor's "Cognitive Science and the Twin-Earth Problem," *Notre Dame Journal of Formal Logic,* XXXII, 2 (April 1982): 98–188, together with Burge's rejoinder "Two Thought Experiments Reviewed," *ibid.,* XXII, 3 (July 1982): 284–293.

be bilingual in English and vat-English [in which case the sentence mentioned in (T) could be in a different language from the language of the words used in (T)]. In the light of these remarks, recall argument E's premise:

(5) If I am a non-BIV (speaking English), then my utterances of 'I am a BIV' are true iff I am a BIV.

Given that (T) would also be true as uttered by a BIV, we apparently also have

(8) If I am a BIV (speaking vat-English), then my utterances of 'I am a BIV' are true iff I am a BIV.

Compare (8) with

(9) If I am speaking a language in which 'tail' refers to legs, then horses have four tails.

But the truth of (8) raises problems for the reconstructed argument. Given premise (3) in E:

(3) If I am a BIV (speaking vat-English), then I do not have sense impressions as of being a BIV.

it follows that (8) and premise (2) give incompatible truth conditions for my utterances of 'I am a BIV' on condition that I am a BIV. The truth conditions specified in

(2) If I am a BIV (speaking vat-English), then my utterances of 'I am a BIV' are true iff I have sense impressions as of being a BIV.

are incompatible with those given in (8) if premise (3) is true. Premise (2), then, seems problematic if (8) is true.

One might respond to the foregoing objection by maintaining that (8) is false for the same reason that (9) is false. That is, one could reasonably maintain that in evaluating the truth value of (9) at a world, we are *not* to (I) interpret its consequent in such a way that the condition specified in its antecedent is taken as applying to the language used in stating the consequent and (II) evaluate the latter's truth value accordingly. It is only if we did so interpret the consequent that it would come out true (holding the rest of English fixed) whenever the antecedent is true. If we did not interpret the consequent in this nonstandard manner, we would evaluate (9) at a world in which its antecedent is true (a world in which a variant of English is spoken) by determining whether in such a world, the *English* sentence 'Horses have four tails' would be true. On this view, (8) is no more plausible than (9) and hence presents no problem for (2). (8)

seemed plausible only because it was mistakenly assumed that, since (T) would be true as uttered by a BIV speaking vat-English, (8) is thereby established. But this fact about (T) establishes (8) only if, in evaluating the truth value of (8) at a world in which its antecedent is true, we must interpret its consequent as being in vat-English and evaluate its truth value accordingly. This method of interpretation, however, is just as nonstandard for (8) as for (9).[14]

This response, of course, depends upon the assumption that, just as in (9) we interpret the consequent as being in English even when we are evaluating its truth value at a world in which a variant of English is spoken, so in (8) we interpret the consequent as being in English when we are evaluating its truth value at a world in which vat-English is spoken—we interpret the consequent of (8) as being an English specification of the truth conditions that the mentioned sentence would have as uttered in vat-English. The idea is that the consequent of (2) gives the correct English specification of those truth conditions. The consequent of (8), *if understood as a piece of vat-English,* would give the correct specification of the mentioned sentence's vat-English truth conditions. But the language used throughout (8), we assume, is English. On this understanding, (8)'s consequent does *not* give the truth conditions that 'I am a BIV' would have as uttered by a BIV. (8) is false, then, because there are worlds at which its antecedent is true, yet its consequent false: worlds in which I am a BIV.

Now this is not an entirely legitimate response to the original objection. If I am allowed to assume that I am speaking English rather than vat-English, then I am allowed to assume that I am not a BIV. In that case, the argument E is of no interest. If I do not assume that the argument is being given in English, though, the problem of evaluating the argument becomes quite bizarre. Normally, when one evaluates an argument, one is allowed to assume that it is stated in a given language L, but this is not so in the case at hand. Consider the following question. Suppose a BIV were to consider the sentences comprising E [(1)–(7)]. Would he then be entertaining a set of propositions that constituted a sound argument? If not, then a problem for the current anti-Cartesian strategy would arise as follows. On Putnam's view (as interpreted in the current anti-skeptical strategy), there is a sense in which I do not know which propositions are

[14] There is surely no temptation to suppose that this variant on (9) is true:

(9′) If I were speaking a language in which 'tail' refers to legs, then horses would have four tails.

(8) and (9), I am claiming in the text, are no more plausible than (9′).

expressed by the sentences in E when I utter them. Which proposition is expressed by my utterances of, say, 'I am a BIV' depends upon whether or not I am a BIV speaking vat-English, and I cannot claim to know that I am not a BIV until I can claim to know that E is a sound argument and that it somehow allows me to know that I am not a BIV. The tactic of the current strategy is to show, by means of E, that no matter which proposition is expressed by 'I am a BIV' when I utter it, what is expressed is *false*. But a condition for the success of the strategy is that I can at least claim to know that the sentences in E express propositions forming a sound argument when I utter them. Can I claim to know this without assuming that I am speaking English?

Before attempting to answer this question, let me explain why our most recent considerations show that the Burge-style anti-skeptical strategy (as I have construed it) is problematic. If I do not know whether I am speaking English or the "language" spoken by a BIV, and if the utterances of a BIV lack content (as Burge seemed to claim in the case of the solipsistic twin), then I do not know whether or not my own utterances have content. I think that they obviously do have content; yet I do not know this unless I know whether or not I am a BIV. So the Burge-style strategy engenders a skeptical problem about the very meaningfulness of my sentences. By contrast, on the reconstructed Putnamian view, my utterances have a determinate content even if I am a BIV speaking vat-English, and even in such a situation these utterances express *truths* (contrary to what the skeptic had thought). One's lack of knowledge as to whether one is speaking English or vat-English also raises a problem for yet a third strategy suggested by Putnam (which is distinct from both his suggested move on which a BIV means nothing by 'I am a BIV' and from the move embodied in E). Putnam at one point says, "although the people in . . . [the vat-world] can think and 'say' any words we can think or say, they cannot (I claim) *refer* to what we can refer to. In particular, they cannot think or say that they are brains in a vat (*even by thinking 'we are brains in a vat'*)" (8). This remark might seem to contain the seeds of an anti-skeptical argument which is different from both the Burge-style "no-content" argument and the argument developed at some length above. Thus, Michael Williams (in a review of *Reason, Truth and History*[15]) says that Putnam adopts a "view of meaning and reference which precludes his taking seriously the skeptical problems which bedevil the metaphysical realist" (260), and Williams goes on to characterize Putnam's anti-skeptical reasoning as follows:

[15] This JOURNAL, LXXXI, 5 (May 1984): 257–261.

"the words or thought-signs used by brains-in-vats, including 'brain', 'vat', etc., do not mean what they mean when used by normal human beings. Thus anyone who can think to himself, 'I may be a brain in a vat,' meaning by this what we would normally mean by it, is not a brain in a vat" (261). However, I can conclude from this that I am a normal human being rather than a BIV—and thereby lay the skeptical problem to rest—only if I can assume that I mean by 'I may be a BIV' what normal human beings mean by it. But I am entitled to that assumption only if I am entitled to assume that I am a normal human being speaking English rather than a BIV speaking vat-English. This must be *shown* by an anti-skeptical argument, not assumed in advance.

IV

To return to our evaluation of the argument E, do the constituent sentences of E express propositions forming a sound argument if these sentences are in vat-English? In order to answer this question, I will try to express the propositions that would be expressed by the sentences in E as uttered by a BIV speaking vat-English. The point will then be to see whether the propositions expressed form a sound argument. Before we can begin to undertake this maneuver, the question arises, what would a BIV's token of 'sense impression' refer to? It seems absurd to say that just as his token of 'tree' would refer to sense impressions as of trees [on account (i)], so his token of 'sense impression' would refer to sense impressions as of sense impressions. It would probably be best for Putnam to say that a BIV's token of 'sense impression' would refer to sense impressions.[16] On this construal, then, according to account (i), a BIV's token of 'sense impression as of trees' would refer to sense impressions as of trees. Here, then, is the set E′ of sentences which (apparently) express the propositions that a BIV would be considering were he to consider the sentences in E [I use account (i) throughout]:[17]

[16] Alternatively: just as 'Here is a tree' has as its vat-English truth conditions the would-be phenomenalist truth conditions for the English reading of the sentence 'Here is a tree', so 'It merely seems as if there is a tree here' has as its vat-English truth conditions the would-be phenomenalist truth conditions of that English sentence. These latter truth conditions, though, are apparently the same as the normal truth conditions of the English sentence about seeming. I am indebted to Michael Thompson here.

[17] On account (ii), the antecedent of the counterpart to premise (2) would be 'I have electrical impulses which cause me to have sense impressions as of being a BIV'. So we would have to ask what a BIV's token of 'electrical impulses which cause me to have sense impressions as of being a BIV' refers to. Presumably, it would refer to electrical impulses which cause me to have sense impressions as of electrical impulses which cause me to have sense impressions as of being a BIV.

(1′) Either I have sense impressions as of being a BIV or I do not have sense impressions as of being a BIV.
(2′) If I have sense impressions as of being a BIV, then my utterances of 'I am a BIV' are true iff I have sense impressions as of being a BIV.
(3′) If I have sense impressions as of being a BIV, then I do not have sense impressions as of being a BIV.
(4′) If I have sense impressions as of being a BIV, then my utterances of 'I am a BIV' are false. [(2′), (3′)]
(5′) If I do not have sense impressions as of being a BIV, then my utterances of 'I am a BIV' are true iff I have sense impressions as of being a BIV.
(6′) If I do not have sense impressions as of being a BIV, then my utterances of 'I am a BIV' are false. [(5′)]
(7′) My utterances of 'I am a BIV' are false. [(1′), (4′), (6′)]

Call the set of sentences (1′)–(7′) *E′*. Does E′ express a sound argument? The argument is valid, and the question whether it is sound hinges upon the truth values of (2′), (3′), and (5′). Whether these premises are true depends upon whether any sense impression could possibly count as a *seeming to be a BIV*, in the way in which my current sense impression clearly counts as a *seeming to be in a library room.* A sense impression as of seeing a room containing a brain in a vat connected to a computer would not count as the former sort of seeming, even if the seeming to see were as if from a disembodied point of view. If such a sense impression would not count as a seeming to be a BIV, then it is hard to imagine which sort of sense impression *would*. If no sense impression could count as a seeming to be a BIV, then Putnam can maintain that (2′) and (3′) are both true by virtue of having a common necessarily false antecedent.[18]

What about (5′)? Its antecedent is not a necessary falsehood. Further, this premise will be false if the conditionals in the argument are read as being stronger than material conditionals (as we have in fact been reading them). This is because even the subjunctive reading of (5′) is false, since if I were a non-BIV who does not have sense impressions as of being a BIV, then the truth conditions for my utterances of 'I am a BIV' would *not* be those specified in (5′)'s consequent. Rather, they would be those specified in (5)'s consequent:

(5) If I am a non-BIV (speaking English), then my utterances of 'I am a BIV' are true iff I am a BIV.

[18] The same point would hold if either account (ii) or (iii) were used in the foregoing argument. For example, on account (ii), the antecedent shared by the counterparts to (2′) and (3′) would apparently be necessarily false (see fn 17).

However, even though the semantical premises of the original argument E were presumably supposed to follow from conceptual facts of some kind concerning reference (and hence the argument's conditionals were thought to be stronger than material conditionals), we need not insist on interpreting the conditionals in E′ as being stronger than material conditionals. So long as the sentences in E′ express a sound deductive argument of some kind regardless of whether they are interpreted as being in English or in vat-English, I can reason that the sentences in the original argument E express a sound deductive argument regardless of whether they are interpreted as being in English or in vat-English. The fact that the vat-English argument would be unsound if the conditionals were interpreted as subjunctive or strict conditionals seems irrelevant so long as the material conditional reading yields a sound argument.

(5′) is true when read as a material conditional and evaluated at a world in which its utterer is a BIV. In such a world, its antecedent is true because a BIV does not have sense impressions as of being a BIV. In a vat world, its consequent comes out true as well, since it is a correct specification (on Putnam's views about reference) of the vat-English truth conditions for 'I am a BIV'. Hence, when a BIV utters sentence (5) of the original set of sentences E, the sentence, read as a material conditional, expresses a true proposition. The same holds, we have seen, for the other premise sentences (since, on their vat-English readings, their common antecedent expresses a falsehood—indeed an apparently necessary falsehood). Further, the argument expressed is valid. So when a BIV utters the sentences in E, these sentences express propositions forming a sound argument (i.e., the sentences express the propositions expressed by the sentences in E′, and these propositions, I have just argued, form a sound argument).[19] On the other hand, when an English-speaking non-BIV utters the sentences in E, we have seen, these sentences also express propositions forming a sound argument.[20]

[19] We are now in a position to see that, for both the BIV and the non-BIV, (2) and (8) are *compatible* so long as they are read as material conditionals. This is because (2) and (8) share a false antecedent on both their English and vat-English readings. When (2) and (8) are read as involving subjunctive or strict conditionals, they are incompatible, but (8) is false (I argued in the text) on its English reading.

[20] I have just reasoned that, no matter whether the sentences (1)–(7) expressing our anti-skeptical argument are in English or in vat-English, the sentences express propositions forming a sound argument. This reasoning was conducted in the metametalanguage in which the relevant part of the text of this paper is stated, viz., the language used to discuss such metalinguistic sentences as (1)–(7). It might be objected that, in the course of that reasoning, I have tacitly assumed that this metametalanguage is English. In particular, one might object that I have in effect been arguing as follows:

The argument expressed by E (whether in English or in vat-English) does not directly refute premise (F) of the Cartesian skeptical argument:

(F) I do not know that I am not a BIV.

The problem is whether I can argue from my knowledge of E's conclusion:

(7) My utterances of 'I am a BIV' are false.

to the further conclusion that I *do* know that I am not a BIV. Let us recall that, whether or not I am a BIV, the following sentence is true as uttered by me:

(T) My utterances of 'I am a BIV' are true iff I am a BIV.

(7) and (T) entail

(10) It is not the case that I am a BIV.

(x) The metalanguage in which the sentences in E [(1)–(7)] are stated is either English or vat-English.

(y) If the metalanguage is English, then the sentences in E express propositions forming a sound argument, given that the sentences in E are to be read homophonically.

(z) If the metalanguage is vat-English, then the sentences in E express propositions forming a sound argument, given that the sentences in E are not to be read homophonically but rather as expressing the propositions expressed by the sentences in E'.

(w) The sentences in E express propositions forming a sound argument. [(x), (y), (z)]

Premises (y) and (z) depend upon the assumption that the metametalanguage in which the relevant part of this paper is stated is English. We need to ask, though, whether the reasoning that supports (y) and (z) is correct if the metametalanguage in which it is conducted is vat-English rather than English. For example, a defense of (y) would require the claim that if a non-BIV uttered

(5) If I am a non-BIV (speaking English), then my utterances of 'I am a BIV' are true iff I am a BIV.

then his sentence would express a true proposition, since it would be read homophonically. This is the claim that:

(CL) If one is a non-BIV, then the proposition that: *if one is a non-BIV, then one's utterances of 'I am a BIV' are true iff one is a BIV* is true.

Suppose that I am in fact speaking vat-English. Then we need to ask whether (CL), as uttered by me, expresses a true proposition. If it expresses a false proposition, then my reasoning in the text on the matter of our anti-skeptical argument's soundness rests upon a false claim. I leave it to the patient reader to verify that (CL) would express a true proposition as uttered by a BIV and that, more generally, the evaluation of our anti-skeptical argument's soundness is unproblematic, even given the assumption that the metametalanguage is vat-English.

Since I know that my utterances of 'I am a BIV' are false and that these utterances are true iff I am a BIV [I know these things on the basis of the argument of this paper and my knowledge of the disquotational principle (T)], it follows that I know that I am not a BIV.[21] Hence, premise (F) of the skeptical argument is false.[22]

V

Given the presuppositions of our anti-skeptical argument, it is difficult to avoid an uneasy feeling that there is some trick involved in the reasoning of the last paragraph. To see that there *is* a trick and that there is thus a severe limit to the anti-skeptical force of our argument, note that we move from that argument's conclusion—(7)—to (10) only by invoking (T). But we have already seen that disquotational principles like (T) must be used quite carefully in contexts such as this. For example, if I do not know whether *S* is speaking English or vat-English, then I cannot apply a disquotational principle analogous to (T) to *S*'s utterances of '*S* is a BIV' and conclude that those utterances are true iff *S* is a BIV. Similarly, if I do not know whether *I* am speaking English or vat-English, then I cannot apply (T) to my own utterances of 'I am a BIV' as a step toward the conclusion that I know that I am not a BIV and hence am speaking English. Another way to see the point is to note that since I do not know whether I am speaking English or vat-English, I do not know whether the truth conditions of my utterances of 'I am a BIV' are the strange ones specified in premise (2) or rather the disquotational ones specified in

[21] This inference requires that knowledge is closed under known logical implication, which we are assuming to be true for the purposes of the present paper.

[22] In "Putnam's Brains" [*Analysis*, XLIV.2, 202 (March 1984): 59–61], Jane McIntyre interprets Putnam's argument as follows:

(1) If we are brains in a vat, then the sentence 'we are brains in a vat' says something false.
(2) If the sentence 'we are brains in a vat' says something false, then we are not brains in a vat.
(3) Therefore, if we are brains in a vat, then we are not brains in a vat (from (1) and (2)).
(4) Therefore, we are not brains in a vat (from (3)).

McIntyre criticizes this reconstruction of Putnam's argument (p. 60) by saying that in premise (1) the purported falsity of the mentioned sentence "derives from its status as a sentence of vat-English, and its . . . [consequent] reference to brain images and vat images." However, she argues, "the falsity of 'I am a brain in a vat' in *vat-English* does not entail that I am not a brain in a vat." Thus, "the argument that supports premise (1) . . . defeats premise (2) and Putnam's argument is therefore unsound." The text's discussion of my own reconstructed argument shows that McIntyre's reconstruction falls short through lack of consideration of the *English* truth conditions for 'I am a brain in a vat'. Such a consideration leads one to propound an argument—E—which is more complex than McIntyre's and which is sound (I have argued) no matter which language it is expressed in.

premise (5) [those given by (T)]. Even on this assumption, our argument ran, I can still know that my utterances of 'I am a BIV' are false. Using (T) in aid of our anti-skeptical argument is thus inconsistent with one of the assumptions behind the argument. And if that assumption is dropped (if I claim to know that I am speaking English rather than vat-English), then there is no longer any need for the anti-skeptical argument.

To see even more clearly that our anti-skeptical argument grinds to a halt at (7)—at the level of sentences, not propositions—consider a case in which a trustworthy set theorist tells me that the sentence 'Omega is not a regular cardinal' expresses a true proposition. Knowing very little about set theory, I do not understand the technical terminology in the sentence. So, even though I can claim to know that the sentence expresses a true proposition, I do not know *which* proposition. Hence I cannot claim to know *that omega is not a regular cardinal,* given only my metalinguistic knowledge that the relevant sentence is true. Suppose you say that surely I know that 'Omega is not a regular cardinal' is true iff omega is not a regular cardinal, and that I therefore know that omega is not a regular cardinal, given my knowledge that the relevant sentence is true. However, isn't it rather that I know only that the following sentence is true: " 'Omega is not a regular cardinal' is true iff omega is not a regular cardinal?" This seems to be what I know in virtue of my knowledge of what 'true' means, given the fact that I do not (by hypothesis) understand the sentence 'Omega is not a regular cardinal'. Hence I cannot claim to know that omega is not a regular cardinal, given only my metalinguistic knowledge that these two sentences are true: 'Omega is not a regular cardinal', " 'Omega is not a regular cardinal' is true iff omega is not a regular cardinal."

The current problem facing our anti-skeptical argument, then, is that is at best affords knowledge that a certain sentence expresses a false proposition, whereas the intended sort of refutation of skepticism depends upon the availability of knowledge that a certain proposition is false—the proposition that I am a BIV. It is useful to compare the result of our Putnamian anti-skeptical argument with the anti-skeptical result afforded by verificationism. In each case, the skeptic is reduced to silence: he cannot succeed in asserting any proposition that is a counterpossibility to our ordinary knowledge claims and not known by us to be false. When the skeptic tries to assert such a proposition, according to the two strategies, he inevitably ends up asserting some *other* unproblematic proposition which *is* known by us to be false. A difference between the two strategies is that verificationism, but not our Putnamian strategy, has it that there

is *no* genuine proposition that is (1) susceptible of confirmation (or infirmation) by sensory evidence, (2) a counterpossibility to our ordinary knowledge claims, and (3) not known by us to be false. This is not a contention of the reconstructed Putnamian strategy, and it might seem that this strategy is therefore preferable to verificationism (since it might seem implausible to hold that the skeptic has not even succeeded in stating an epistemological problem that is at least superficially interesting). According to the Putnamian strategy, there *is* a genuine counterpossibility proposition such that (i) if I am not a BIV, then my sentence 'I am a BIV' expresses that proposition, and (ii) if I am a BIV, then even though the proposition is true, it neither is expressed by my sentence 'I am a BIV' nor constitutes a counterpossibility to anything I claim to know (any proposition expressed by a sentence I assent to). The skeptic's problematic proposition exists, according to the reconstructed Putnamian strategy, yet is not what a BIV knows to be false when he knows that the proposition expressed by his sentence 'I am a BIV' is false. A BIV knows some *other* proposition to be false, viz., the proposition expressed by his utterances of 'I am a BIV'.

The fact that our reconstructed anti-skeptical strategy, unlike verificationism, countenances the skeptic's supposedly problematic proposition does not in the end constitute a ground for preferring our strategy to verificationism. Countenancing the apparently genuine counterpossibility proposition is one thing, but it is quite another to go on to acknowledge that even though I know that my sentence 'I am a BIV' expresses a false proposition, I do not know whether or not it expresses the skeptic's counterpossibility proposition. If I do not know whether or not the sentence expresses that problematic proposition, then our anti-skeptical argument has not enabled me to conclude that I know that I am not a BIV.

The anti-skeptical strategy reconstructed herein fails in the end because it engenders a sort of skepticism about meaning or propositional content. According to the presuppositions of this strategy, the sentence 'I am a BIV' has different truth conditions in vat-English from those it has in English, and therefore the sentence expresses a different proposition in vat-English from that which it expresses in English. So if I do not know whether I am speaking vat-English or English, then I do not know which proposition my utterance of 'I am a BIV' expresses.[23] Though this kind of lack of knowledge (I have

[23] Whereas the Putnamian anti-skeptical strategy is severely limited by the fact that it engenders skepticism about what my sentences mean, the Burge-style anti-skeptical strategy, as noted above, is severely limited by the fact that it engenders skepticism about whether my sentences mean anything at all (about whether they express any determinate propositional content at all).

argued) does not affect my claim to know that our anti-skeptical argument is sound, it clearly does place a limit on what I can claim to know by virtue of knowing that argument's conclusion. All I can claim is the metalinguistic knowledge that a certain sentence expresses a false proposition, rather than the object-language knowledge that I am not a brain in a vat.[24] Since the latter knowledge was required in order to refute the skeptical argument in the envisaged manner, the present anti-skeptical strategy fails.[25]

ANTHONY L. BRUECKNER

Yale University

[24] One might hold that, even if 'I am a BIV' has different truth conditions in vat-English from those it has in English, it does not follow that there is a difference in propositional content between the beliefs expressed by vat-English and English utterances of the sentence. Fodor, for example, would hold that, despite the difference in truth conditions, a vat-English speaker and an English speaker who both uttered 'I am a BIV' should both be ascribed the same (de dicto) belief content (this would be his position according to "Cognitive Science and the Twin-Earth Problem," motivated by his methodological solipsism with respect to cognitive science). Though such a view would block the objection in the text to the Putnamian argument, it would at the same time contradict the main assumption behind the argument: in virtue of the difference in (causal-theoretically determined) truth conditions, vat-English and English utterers of 'I am a BIV' believe different propositions, and in each case a *true* proposition is believed.

[25] There is a strong similarity between the Putnamian anti-skeptical strategy discussed here and that espoused by Paul Horwich in "How to Choose between Empirically Indistinguishable Theories," this JOURNAL, LXXIX, 2 (February 1982): 61–77 (see especially pp. 75/6; Hartry Field called my attention to this similarity). The problems discussed in the text, I would argue, equally afflict Horwich's strategy. It is worth noting that the formulation of our normal theory of the external world and the formulation of the BIV theory are not *potential notational variants* of the sort Horwich discusses in arguing for his claim that we can know a priori that a potential notational variant S_2 of "the actual formulation of our beliefs" S_1 is false. In that discussion (see p. 66), it is crucial that a potential notational variant S_2 of S_1 will differ from S_1 only in the following way: S_2, when interpreted homophonically, says that, e.g., some entities other than *trees* satisfy all the beliefs expressed in S_1 using 'tree'. In that case, Horwich claims, we know a priori that S_2 (interpreted homophonically) is false. This is because, if anything satisfies all the aforementioned beliefs, trees do. This reasoning, though, cannot be straightforwardly extended to the BIV case, since a formulation of the BIV theory will not say that some entities other than trees are green, leafy, etc. The BIV theory's formulation, when interpreted homophonically, will say that *nothing* is leafy, green, etc.