

Introduction

Anthony Brueckner and Gary Ebbs

Language users ordinarily suppose that they know without empirical investigation what thoughts their own utterances express. Call this supposed knowledge *minimal self-knowledge*. What does it come to? And do we actually have it? These questions are puzzling on just about any view of meaning, reference, and the nature of mental states. They are puzzling in a special way, however, if one accepts *anti-individualism*, which implies that the thoughts that a person's utterances express are partly determined by facts about her social and physical environments. The problem is that if anti-individualism is true, then there are some apparently intelligible skeptical hypotheses that threaten to undermine our ordinary supposition that we have minimal self-knowledge.

In the essays collected in this volume, we debate how to characterize this problem and what it shows. One of us (Brueckner) argues that in some skeptical contexts, there is a coherent, powerful, *prima facie* worry that we lack minimal self-knowledge, given the assumption of anti-individualism, while the other (Ebbs) argues that the apparent intelligibility of the allegedly problematic skeptical hypotheses is illusory, and hence does not undermine our supposed minimal self-knowledge. In this introduction we sketch some of the essential background for our debate and present a brief overview of our essays.

I ARGUMENTS FOR ANTI-INDIVIDUALISM

The following argument illustrates a general pattern of arguments for anti-individualism (taken from Putnam 1975 and Burge 1979) that we presuppose in our debate:

Step one

We imagine that Oscar, an ordinary English speaker who is competent in the use of the English word 'water' but does not accept (or reject) the

sentence ‘Water is H_2O ’, utters a sentence containing the word ‘water’, for instance, the sentence ‘Water is a liquid at room temperature’. Since Oscar is a competent English speaker, other English speakers take his word ‘water’ to be the same as their word ‘water’, hence they take him to have said that *water is a liquid at room temperature*. If, in addition, they think his utterance is sincere, they take him to believe this.

Step two

We stipulate that there is a planet called Twin Earth which is just like Earth except that wherever there is water on Earth there is twin-water, a liquid with an underlying chemical structure that is very different from the chemical structure of water, on Twin Earth. We suppose that on Twin Earth there lives a person we may call Twin-Oscar, who is a physical, phenomenological, and behavioral twin of Oscar. Twin-Oscar is a normal speaker of Twin-English, the Twin-Earth counterpart of English. When Twin Oscar utters the sentence ‘Water is a liquid at room temperature’, his fellow Twin-English speakers take his word ‘water’ to be the same as their word ‘water’, hence they take him to have said (when translated into English) that *twin-water is a liquid at room temperature*. If, in addition, they (and we) think his utterance is sincere, they (and we) take him to believe this.

These two steps together support anti-individualism. For, by hypothesis, Oscar and Twin-Oscar share all the same *individualistic properties*: those which concern their qualitative perceptual experience and stream of consciousness, their behavior and behavioral dispositions, and their functional states. Even so, Oscar’s word ‘water’ differs in reference from Twin-Oscar’s ‘water’, which correctly applies to all and only samples of XYZ, rather than to all and only samples of H_2O . Oscar’s word ‘water’ and Twin-Oscar’s word ‘water’ thus have different extensions and express different concepts. The truth conditions of Oscar’s ‘water’-sentences accordingly differ from those of Twin-Oscar’s corresponding ‘water’-sentences. When Oscar utters his sentence ‘Water is a liquid at room temperature’, he thinks *that water is a liquid at room temperature*, whereas Twin-Oscar, who simultaneously utters his sentence ‘Water is a liquid at room temperature’, thinks (as we express it) *that twin-water is a liquid at room temperature*. If we suppose that the content of a person’s mental state is what is specified by what follows ‘that’ in a description of that mental state, then the thought that Oscar thinks by using his sentence ‘Water is a liquid at room temperature’ differs

in *content* from Twin-Oscar's corresponding thought. These differences in content apparently derive from the differences between Oscar's and Twin-Oscar's external social and, especially, physical environments – Oscar's replete with water (H₂O), and Twin-Oscar's replete with twin-water (XYZ). The observations therefore support anti-individualism, according to which a thinker's external physical and social environment partly determines the semantic properties of his words and sentences, and the contents of his intentional mental states. (Following standard usage, we sometimes call anti-individualism about content *content externalism*, and anti-individualism about semantic properties, such as reference, *semantic externalism*.)

2 MENTAL CONTENT AND INCOMPLETE UNDERSTANDING

Gareth Evans (Evans 1982, Chapter II) grants that speakers of a natural language typically take each other's utterances at face value in the way that is illustrated by the Twin-Earth thought experiments, but insists that a speaker cannot be credited with having a mental state with a given content unless he has accurate and complete beliefs about what that content is. On Evans's view, a speaker may utter a sentence that expresses a content which is not the content of any of the speaker's own mental states. In the situation described above, for instance, when Oscar utters the sentence 'Water is a liquid at room temperature', he thereby expresses the content *that water is liquid at room temperature*, but, according to Evans, he does not *also* thereby express or have any mental state, including any thought, with that content. Evans's view implies that in the cases that we focus on in this volume, speakers do not express thoughts whose contents match the contents of the public language sentences they utter, and hence the question about minimal self-knowledge that concerns us in this volume – the question of how a person can know without empirical investigation what thoughts her utterances express, assuming that the contents of her thoughts are the same as the contents of the sentences she utters – does not arise.

Unlike Evans, we take arguments for anti-individualism to show that in ordinary situations, at least, when skepticism is not in question, there is no distance between a speaker's sincere utterance of a sentence that expresses a particular content, such as the content that water is a liquid at room temperature, on the one hand, and the speaker's thereby expressing a mental state of his with that content, on the other. We take arguments for anti-individualism to show that even speakers who do not have a full understanding of the concepts expressed by the words that they utter ordinarily possess mental states with contents that contain those concepts.

Assuming anti-individualism, understood in this way, our debate focuses on the question of how one can take oneself to have minimal self-knowledge in contexts in which one entertains certain radical skeptical hypotheses.

3 OVERVIEW OF THE CHAPTERS

Chapters 1–13 fall into three main groups, comprising Chapters 1–5, 6–11, and 12–13, respectively. In Chapters 1–5, we focus on questions about minimal self-knowledge that are raised by Putnam’s argument that we are not brains in a vat, and we come to see that a central question for us is whether doubts about one’s apparent knowledge of the contents of one’s own mental states are coherent. In Chapters 6–11 we debate this question in detail, starting with an argument Ebbs presents in Chapter 6. In Chapters 12 and 13 we each summarize and further develop our different views of what Chapters 1–11 establish and where they lead.

Here is a bit more detail about the chapters in each of these groups.

Chapters 1–5

Chapter 1, Brueckner’s “Brains in a vat,” was the first extended analysis of Putnam’s approach to the problem of skepticism in Chapter 1 of *Reason, Truth and History* (Putnam 1981), where Putnam sought to use the semantic externalist component of anti-individualism to construct an argument that would rule out the skeptical hypothesis that one has always been a massively deceived brain in a vat whose experiences are systematically caused by a complexly programmed supercomputer. The neo-Cartesian skeptic maintains that one does not know, for example, that one has hands, in virtue of one’s inability to knowledgeably rule out the vat hypothesis. Putnam’s starting point was the idea that due to the differences between the causal environments in which normal thinkers and brains in vats are ensconced, the semantic properties of the language of the brain in a vat (supposing that it thinks in a language) differ from those of the language spoken by his normal unenvatted counterpart. Putnam thought that this difference would enable one to argue that one is *not* a handless brain in a vat. In Chapter 1, Brueckner reconstructs an anti-skeptical argument from these Putnamian materials. Brueckner’s main worry about the reconstructed argument is that given the dialectical situation between the skeptic and the Putnamian anti-skeptic, the semantic externalism that drives the argument also undercuts the argument. The worry is that if at the outset of the anti-skeptical argument, one does not know whether one is a normal

human or, instead, a brain in a vat, then one does not know the meanings and associated truth conditions of one's sentences. If this worry is well-founded, then Putnam's argument does not enable one to establish that one is not a brain in a vat. The worry presupposes that there can be coherent skeptical doubts about one's apparent knowledge of the semantic properties of one's own language.

Chapter 2, Ebbs's "Skepticism, objectivity and brains in vats," argues, among other things, that such doubts, as well as corresponding doubts about one's apparent knowledge of the contents of one's own mental states, are incoherent. Ebbs argues that to understand and evaluate Putnam's argument that we are not always brains in vats, we must not simply grant the skeptic from the start that *we may actually be brains in vats from the beginning to the end of our lives*, as Brueckner does in his assessment of his reconstructed Putnamian argument. Instead, we must take seriously from the start that we might not be able to make sense of actually being brains in vats from the beginning to the end of our lives. Ebbs argues, in effect, that if we regard it as a genuine possibility that we cannot make sense of actually being brains in vats, then the main results of anti-individualism support Putnam's reasoning. Ebbs also aims to rehabilitate the Putnamian considerations by shifting from the use of semantic externalism to the use of anti-individualism about mental content in the construction of an anti-skeptical argument. Applying these two aspects of his reading of Putnam's argument, Ebbs argues that Thomas Nagel's well-known response to Putnam in *The View from Nowhere* (Nagel 1986) depends on a misunderstanding of anti-individualistic accounts of belief content.

Chapter 3, Brueckner's paper "Ebbs on skepticism, objectivity and brains in vats," responds to Ebbs's criticisms in Chapter 2, holding that an analogue to Brueckner's worry about his own semantic externalist argument arises for the Ebbs-style argument that is fueled by content externalism. Assuming again, as he does in Chapter 1, that there can be coherent skeptical doubts about one's apparent knowledge of the semantic properties of one's own language, Brueckner suggests that in the dialectical situation in play in the skeptic/anti-skeptic dispute, claims to know the contents of one's own thoughts are just as problematic as claims to know the semantic properties of one's own language. Brueckner also discusses the question of whether one can extend the anti-individualist approach to non-empirical concepts, a question that Ebbs raises in his discussion of Nagelian objectivity in Chapter 2.

In Chapter 4, a new essay written for this volume, Ebbs responds to Brueckner's criticisms in Chapter 3 by distinguishing between two different

conceptions of the dialectical context of Putnam's argument. Ebbs grants that Putnam's argument must begin with an attitude of agnosticism about whether we are in a vat world or a normal world, but he does not accept Brueckner's understanding of the sense of agnosticism that is relevant to the dialectical context of Putnam's argument. Brueckner's understanding of the agnosticism rules out an important kind of response to a skeptical argument, one that begins by entertaining the possibility that a given skeptical hypothesis is coherent, and may actually be true, and ends by concluding that the hypothesis, while meaningful, is not coherent, and therefore cannot actually be true. According to Ebbs, it is this latter kind of response to skepticism that is relevant to the dialectical contexts of Putnam's argument and of Ebbs's related criticism in Chapter 2 of Thomas Nagel's conception of objectivity.

Chapters 2–4 highlight the central importance for both Brueckner and Ebbs of the question whether doubts about one's apparent knowledge of the contents of one's own mental states are incoherent. Chapter 5, Brueckner's "Trying to get outside your own skin," canvasses and criticizes several of the ways in which philosophers, including Ebbs, have tried to respond to the charge that externalist semantic views generate a coherent, powerful, *prima facie* worry that we lack minimal self-knowledge.

Chapters 6–11

The second main group of chapters begins with Chapter 6, Ebbs's "Can we take our words at face value?" which presents a new way of responding to the charge that externalist semantic views generate a coherent skeptical challenge to our ordinary assumption that we have minimal self-knowledge. Ebbs argues that we can't even try to raise a skeptical challenge to our ordinary beliefs about what thoughts our utterances express unless we take for granted that we are competent to use our words to raise the challenge, and this requires that we presuppose some background empirical beliefs. But if we presuppose background empirical beliefs that are sufficient for minimal competence, Ebbs argues, then we cannot make sense of the content skeptic's claim that we may actually be in a world in which our utterances express thoughts that are different from what we take them to express.

Chapter 7, Brueckner's paper "Is skepticism about self-knowledge incoherent?" replies to Chapter 6 by attempting to show that we can know we have presented a *sound* argument for content skepticism even if we do not presuppose any empirical beliefs, and hence don't know

what thoughts we express when we utter the premises of the argument for content skepticism. Against this, Chapter 8, Ebbs's "Is skepticism about self-knowledge incoherent?" observes that Brueckner's reply in Chapter 7 depends on the assumption that we are not in any possible world in which the argument that we express by uttering the sentences of the skeptic's argument is unsound. Ebbs argues that the principles of anti-individualism imply that for each of us there are such weird possible worlds. Ebbs also points out that if Brueckner's account of what we can know without empirical investigation is correct, then we cannot know without empirical investigation that we are not in one of the weird possible worlds. Ebbs concludes that Brueckner's reply in Chapter 7 depends on an empirical assumption – the assumption that we are not in one of the weird worlds – that by his own standards we cannot be justified in accepting without empirical investigation. This observation prompted further objections and replies, developed in Chapters 8–11, which concern the difficult questions of whether one can be justified in accepting the premises of the argument for content skepticism, and, if not, whether that shows that content skepticism is incoherent. In addition to these central issues, Chapters 8–11 range over such related topics as the use of *reductio* arguments in philosophy and skepticism about reasoning itself.

Chapters 12–13

In Chapter 12 and Chapter 13, both of them new and written just for this volume, we each summarize and further develop our own views of what the debate has established and where it leads.

In Chapter 12, "Self-knowledge in doubt," Ebbs lays down three conditions on a successful skeptical argument of the sort that Brueckner aims to construct, and then summarizes and expands on his criticisms of Brueckner's attempts to raise a skeptical doubt about whether we have minimal self-knowledge. According to Ebbs, his criticisms show that we should reject the conception of minimal self-knowledge on which Brueckner's skeptical arguments rely. Ebbs argues that if we adopt instead the minimalist conception of minimal self-knowledge that he sketches in Chapters 6, 8, and 10, we simultaneously avoid the problems with Brueckner's reasoning, and commit ourselves to principles from which it follows that we cannot make sense of doubting that we have minimal self-knowledge. Ebbs ends the chapter by sketching a methodological framework within which our failure to be able to raise a coherent doubt about

whether we have minimal self-knowledge helps us to see both what it is and that we have it.

In Chapter 13, “Looking back,” Brueckner revisits the exchange with Ebbs in the first group of chapters regarding Putnam’s brains in vats, commenting on Ebbs’s new, previously unpublished Chapter 4. Brueckner then tries to summarize the exchange regarding skepticism about knowledge of thought-content embodied in Chapters 6–11, while trying to explicitly formulate principles about a stretch of reasoning’s being *self-undermining*. A discussion of Descartes on the latter issue concludes the chapter.

CHAPTER I

*Brains in a vat**Anthony Brueckner*

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 chapter, 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. 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

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 Putnam 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 Putnam 1978 and Putnam 1980). 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 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 Brueckner 1984.

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 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,

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.

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'.