

Intuitions*

Herman Cappelen & Douglas G. Winblad

Consider the following common philosophical procedure. Some philosophical topic—such as reference, knowledge, or personal identity—is under investigation. A theory is proposed and is then tested against intuitions about relevant hypothetical or actual cases. These intuitions take the form of judgements about whether these cases are cases of a certain kind—cases of reference, knowledge or personal identity, for example. If our intuitions contradict what the theory implies about whether, say, S refers to x, or knows that p, or is identical to T, this generally counts against the theory. If on the other hand, our intuitions match the theory, this counts in its favor. We call this approach to philosophical theorizing the ‘intuition method’. Attempts to specify and evaluate necessary and sufficient conditions for reference, knowledge, personal identity, and the like typically rely on the intuition method. So do more modest attempts to articulate merely necessary conditions in connection with such topics.¹

Why are so many philosophers confident that intuitions about cases are relevant and reliable evidence for and against philosophical theories? The fact that explicit justifications of the intuition method are rare² is nearly as striking as philosophers’ widespread reliance on it. Many of those who use

* Earlier versions of this paper were presented at the Pacific Division meeting of the American Philosophical Association in 1999, the Twentieth World Congress of Philosophy, the University of Oslo, the State University of New York at Albany, and Vassar College. Some of this material appears in “Reference’ Externalized and the Role of Intuitions in Semantic Theory,” forthcoming in *American Philosophical Quarterly*. We thank all those who commented on our work, especially Jennifer Church, Alexander Cappelen, and Daniel Gaskill.

1 For specific examples of discussions that conform to one or the other of these descriptions, see Gettier’s and Nozick’s work in epistemology (Gettier 1963, Nozick 1981), Kripke’s, Putnam’s, and Burge’s accounts of meaning and reference (Kripke 1980, Putnam 1975, Burge 1979), Williams’s and Perry’s treatment of personal identity (Williams, 1973, Perry 1976), and Mackie’s analysis of ethics (Mackie, 1977).

2 See Bealer, 1996, for an interesting, although problematic attempt.

the intuition method think of themselves as engaged in a form of conceptual analysis. The justification they would give for their use of the method accordingly goes something like this:

Philosophers are engaged in the analysis of certain concepts, among them the concepts of knowledge, reference, and personhood. Initially, our grasp of these concepts is largely or wholly tacit. What we are attempting to do as philosophers is make this tacit knowledge explicit. Because intuitions about relevant cases are reliably guided by these tacitly known concepts, such intuitions constitute trustworthy evidence for and against philosophical theories.

We call this rationale for the intuition method the 'concept model'. Many arguments in philosophy can be understood as based on an implicit—or, less frequently, an explicit—acceptance of this model.

Consider, for example, the familiar epistemological enterprise of specifying putative necessary and sufficient conditions for knowing, and evaluating them by determining whether they are compatible with our intuitions about whether specific hypothetical or actual cases are indeed instances of knowing. On the concept model, the epistemologist who engages in this *a priori* enterprise is investigating our tacitly known concept of knowing. The approach is legitimate because our intuitions are reliably guided by our implicit grasp of the concept.³

We shall argue that the concept model fails to provide an adequate justification for the intuition method. In Parts I and II we illustrate an argumentative strategy that is designed to show that the concept model does not legitimize the use of the intuition method with respect to particular philosophical terms. This strategy relies in part on an answer to a largely unexplored—and, we think, intrinsically interesting—question, namely, do externalist arguments apply to philosophical terms? We argue that they do apply to some such terms, and probably apply to all of them. Although our argument only calls into question one attempt to justify the intuition method, we are skeptical about the prospects for alternative justifications. In Part III we give some of our reasons why.

Part I: Externalism regarding Philosophical Terms

Externalism comes in many flavors. We will consider only two of them, Kripke's and Putnam's, although extending our arguments to other variants,

such as Burge's,⁴ is a straightforward matter. Kripke and Putnam have not demarcated precisely the class of terms they believe are subject to externalist treatment,⁵ so it is not clear whether they think their arguments apply to philosophical terms like 'reference', 'knowledge', and 'same person'. In this part we consider the implications for the concept model of assuming that their views extend to philosophical terms. Then, in Part II, we argue that their arguments can be extended in this way.

Kripkean externalism about philosophical terms

According to Kripke,⁶ referential links between names and natural kind terms on the one hand, and the world on the other, are not forged by means of senses, but by initial "baptisms" which are causally related to subsequent uses of the expressions. Since concepts, as understood here, are at the level of senses, Kripkean externalism says that certain terms do not express concepts. It follows trivially that for anyone attracted to a Kripkean externalism with regard to philosophical terms, the concept model is unavailable as a justification for the intuition method. If you are a Kripkean about natural kind terms, you do not think that there is a tacitly known concept of water that gives rise to reliable intuitions about whether particular samples of liquids are samples of water. The same point applies if one is a Kripkean regarding philosophical terms such as 'reference', 'knowledge', 'personal identity', 'freedom', 'time', and so on.

Putnamian externalism about philosophical terms

It is of course compatible with Kripke's view that speakers associate all sorts of information with, say, 'water'. He does not call this information 'the concept of water' because it does not perform the function traditionally associated with concepts. At this point, Kripke's view differs, terminologically, and perhaps even substantively, from Putnam's. One who uses certain terms properly needs to have more in her head on Putnam's view than on Kripke's. In particular, Putnam holds, in order to use a natural kind term meaningfully, one must associate a "stereotype" with the term. He calls the stereotype a 'conventional idea', and sometimes even calls stereotypes 'concepts'. As he puts it, "just such a conventional idea is associated with 'tiger', with 'gold', etc., and, moreover, ... this is the sole element of truth in the 'concept' theory."⁷

4 Burge, 1979.

5 Putnam does write, "Not only does the account given here apply to most nouns, but it also applies to other parts of speech." Putnam, 1975, 244.

6 Kripke, 1980.

7 Putnam, 1975, 250.

3 Of course, someone who thinks about the intuition method in terms of the concept model need not think of philosophical investigations as restricted to concepts. An investigation of a concept can be understood to be an indirect examination of what falls under it as well.

Putnam holds that the stereotype associated with a term “includes” certain properties that are conventionally thought to be possessed by what is in the term’s extension. But it is not necessary that all or most or even most normal things in its extension possess these properties; nor does possessing them ensure that something lies in the extension. Being striped is one of the properties included in the stereotype associated with ‘tiger’, for example, but there can be tigers without stripes. And XYZ possesses all the stereotypical properties of water, but is nevertheless not water.

Note that if our intuitions about C’s are guided by a stereotype that includes properties $P_1 \dots P_n$ and the stereotype leads us to judge that entities with $P_1 \dots P_n$ are C’s, then, since by Putnam’s admission there can be C’s that do not have $P_1 \dots P_n$ and there can be entities with $P_1 \dots P_n$ that are not C’s, our intuitions *will not necessarily* be reliable.

This does not rule out the possibility that some intuitions so guided *are* reliable. It might as a matter of fact be the case that all or most entities with $P_1 \dots P_n$ in our environment are C’s and that there are no—or extremely few—C’s that do not have $P_1 \dots P_n$. This may well be the case for water, tigers, cats and many other natural kinds. If so, our intuitions, guided by the stereotype, will be contingently reliable. One might even argue that it is likely, on evolutionary grounds, that this is in fact the case for a wide range of stereotypes.

But would contingently reliable intuitions make it possible for an externalist to use the concept model in a justification of the intuition method? According to the concept model, philosophers care about intuitions because they are engaged in conceptual analysis. Analysis of a concept C is typically an attempt to come up with necessary and sufficient conditions for being a C.⁸ Contingently reliable intuitions will not enable you to do this. If your intuitions about C’s are reliable because of certain contingent facts, for instance that all and only C’s fit the stereotype of C’s, they will, *at best*, enable you to discover materially adequate conditions, not necessary and sufficient conditions. To see why this is so, suppose someone suggests a set, S, of necessary and sufficient conditions, for being a C. To test S using her intuitions about specific cases, a philosopher would want to consider merely possible cases as well as actual cases.⁹ Now, among the possible situations in which S should be tested are those in which the stereotype does not pick out C’s and those in which there are C’s that do not fit the stereotype. We know there are such situations because the reliability of the stereotypes is contingent. We also know that with regard to such worlds, our intuitions, guided as they are by the stereotype, are *necessarily* misleading as a test for S. So contingently

⁸ Our subsequent remarks also apply, however, to the partial analysis of a concept in terms of merely necessary conditions.

⁹ Merely possible cases, of course, are often the only ones philosophers consider.

reliable intuitions will not help a proponent of the concept model. Since on Putnamian externalism regarding philosophical terms our intuitions are at best contingently reliable, construing the concept model in these terms helps us justify the intuition method as little as does interpreting it along the Kripkean lines described in the preceding section.

Part II: Internalism regarding Philosophical Terms

The foregoing arguments make it clear that the concept model is available as a justification of the intuition method only to someone willing to defend an internalist treatment of philosophical terms. There are two relevant options here: thoroughgoing, or “global” internalism with respect to all terms, and internalism about some terms—including some philosophical terms—coupled with externalism about certain other terms. We call the second sort of position ‘mixed internalism’. There can be various sorts of mixed internalism, depending on which terms are held to lie in the internalist subclass. Some forms of mixed internalism would characterize all philosophical terms in an internalist way. Others would hold that only some such expressions should be understood in this manner. One might for example hold that ‘knowledge’ should be understood in an internalist fashion, but that ‘good’ should not.¹⁰

A mixed internalist needs a justification for limiting the application of externalist arguments to certain classes of expressions. Given that the mixed internalist endorses externalist arguments for other terms, we are pessimistic about whether these requirements can be met. The commitments the mixed internalist undertakes by accepting these arguments make it problematic for her to make an exception of philosophical expressions.

We will illustrate this difficulty by means of two examples. The first extends Putnam’s Twin Earth argument to ‘reference’; the other applies Kripke’s strategy to ‘knowledge’. The point of the examples is not to establish that externalism applies to all philosophical terms, but rather to exemplify a strategy for determining on a case by case basis whether it applies to particular philosophical terms. The ease with which the arguments can be extended in these two examples suggests to us, however, that a large number of philosophical terms may well be susceptible to the same sort of treatment.

¹⁰ There is in the literature a discussion of whether ‘good’ can be subjected to an externalist analysis. Sayre-McCord, 1997, for example, argues that ‘good’ is subject to an externalist treatment, responding in part to arguments formulated by Terrence Horgan and Mark Timmons.

How to extend Putnamian externalism to a philosophical term

In Putnam's Twin Earth thought experiment we are asked to imagine a speaker, *S*, in this world, who uses the word 'water' to refer to water at a time prior to the discovery of the chemical composition of water. We are then asked to imagine a Twin Earth just like Earth except that the chemical composition of the liquid its inhabitants call 'water' is XYZ rather than H₂O. While this is a significant difference, it does not manifest itself in any of the two liquids' "superficial" features. XYZ appears to *S*'s double on Twin Earth, Twin *S*, exactly as H₂O appears to *S*. Given their ignorance of chemistry, the properties *S* and Twin *S* associate with 'water' are therefore the same. Nonetheless, we are inclined to say that when *S* says "I like water," she is referring to H₂O, but when Twin *S* says "I like water" she is referring to XYZ. Because *S* and Twin *S* are in the same psychological states, the extension of 'water' is not determined by such states. Since meaning is what determines extension, Putnam concludes, meanings "just ain't in the head."

Here is a similar argument applied to 'reference'. Suppose Kripke has given a correct account¹¹ of what reference is, one that correctly describes the extension of 'reference' in the actual world.¹² Now imagine a Twin Earth, exactly like ours, except that in it, the extension of 'reference' fits Gareth Evans's theory of reference.¹³ On Earth, what is in the extension of 'reference' are relations between words and the world in which certain expressions are connected to extralinguistic elements by means of causal chains that begin with initial "baptisms" or "dubbings," in accordance with Kripke's account of reference. On Twin Earth, however, what is in the extension of 'reference' is different; typically, the relations in question are

- 11 We leave aside the issue of whether what we have in Kripke is an account, or, rather, merely a "picture." If the reader insists that it is the latter, let him or her suppose for the sake of argument that it has been suitably refined into a theory.
- 12 It is customary to think of a term's extension as the set of things of which the term is true. But, as Putnam notes, it is controversial just what extensions are. See Putnam, 1975, 216–219. We endorse no particular account of the extension of 'reference' here.
- 13 Unlike Kripke, who completely rejects the description theory, Evans supports a version of this theory that incorporates elements of Kripke's causal account. See Evans, 1973. Evans tries to reconstruct from Kripke's "picture" of reference a theory of reference couched in terms of necessary and sufficient conditions, to which he then proceeds to offer alleged counterexamples. It is interesting to note that Kripke does not think his view is in principle incompatible with all of these examples. See Kripke, 1980, 163. We assume here, however, that there is at least one statement of the form 'Expression *e* refers to *o*' with respect to which their accounts differ. If it turns out that none of Evans's putative counterexamples to Kripke's account really is a counterexample, then let us think of the theory which fits the extension of 'reference' on Earth as being one to which these cases do constitute counterexamples.

such that the expressions must be linked to the causal sources of the information associated with them, along Evansian lines.¹⁴ Just as the differences between XYZ and H₂O only manifest themselves in unusual circumstances—such as under laboratory conditions—the differences between the extensions of 'reference' on the two planets are not ordinarily apparent. Both accounts, for example, are compatible with its being the case that the information associated with expressions is ordinarily caused by what they pick out and that expressions are usually introduced by means of baptisms.

Putnam's original thought experiment concerns a period in history prior to the development of modern chemistry. In a similar vein, let us locate our worlds in a time that precedes recent developments in semantics. Now consider Alice, an Earthian who is not a philosopher and does not spend much of her time thinking about semantics. Alice is familiar with the word 'reference'; she uses it occasionally, and by normal standards does so competently. If asked what she associates with 'reference', she says things like "a word's being about something, or standing for something." Such remarks exhaust the information she associates with reference at a general level. She also has idiosyncratic beliefs expressed in sentences such as 'I referred to my sister yesterday', and 'I used "you" to refer to my friend'. Generally, like all of us, when she refers, Alice has no characteristic experience of referring. She may hear or see herself utter or inscribe a word; she may have fleeting images or thoughts associated with its referent. But then again, she may not.

Now consider her double, Twin Alice. When asked what she associates with 'reference', Twin Alice utters the same sentences Alice does. And, phenomenologically, her situation with respect to what she calls 'reference' is indistinguishable from the one Alice is in regarding what she calls 'reference'. The extensions of 'reference' for Alice and Twin Alice differ; but what is in their minds is the same.¹⁵ If this sort of situation shows externalism to be true about 'water' it should do the same for 'reference'.

14 Evans holds that it is "typically necessary" that the referent of a term be the causal source of the information associated with it. In some cases, according to Evans, the referent is merely the *dominant* source of information associated with a term; in others, a speaker who uses a term refers to whatever those to whom he defers refer when they use the term. See Evans, 1973.

15 We are not suggesting, of course, that the aspects of Alice and her double's psychologies that are relevant to their use of 'reference' are exhausted by our descriptions. But we are supposing that whatever else may be involved—unconscious processing, for example—is something with respect to which they are psychologically indistinguishable.

In Putnam's original example, *S* and Twin *S* never encounter evidence of the differences between XYZ and H₂O; they never enter laboratories in which the chemical

How to extend Kripkean externalism to a philosophical term

What about Kripkean externalism? On one construal, the argument in *Naming and Necessity* is remarkably simple. Kripke simply points out that for proper names and natural kind terms, speakers often refer without being in possession of individuating information about that to which they are referring—as in the case he describes of someone who refers to Feynman using ‘Feynman’, but when asked who Feynman is, can only reply, “He’s a physicist, or something.” Sometimes we succeed in referring to something about which we have false information—as in the case in which someone refers to Gödel, even though all the information he associates with the name ‘Gödel’ picks out Schmidt. If this is so, the internalist cannot be right in claiming that in order to use a name to refer to an object, the speaker must possess individuating information about that object. Kripke presents the causal theory as an explanation of how reference can occur in such circumstances—that is, ones in which speakers have little or false information about what they are talking about.

Can analogous arguments be constructed to show that the extensions of philosophical terms are determined in the way the causal theory describes? Given what we have said so far, the answer seems to be yes. Consider Sophia, who is asked what she associates with ‘knowledge’. After thinking about the question for some time, she says, “I’m not sure exactly what knowledge is; it’s a state that requires some sort of justification, but I’m not sure what more to say.” Let us take this answer at face value as telling us what’s “in her mind” when she uses the word. It appears that Sophia can talk about knowledge using the word ‘knowledge’ without possessing individuating information about it, without knowing necessary and sufficient conditions for knowledge.

The philosophical community itself provides us with evidence that one can talk about a philosophical topic while having radically inaccurate information about it, since philosophers have contradictory beliefs about knowledge, personal identity, reference, and so on. If considerations such as compositions of these liquids are being analyzed. Similarly, we suppose in our example that Alice and Twin Alice never encounter evidence of the differences between the extensions of ‘reference’ on their home planets. Moreover these differences need not be such that Alice and Twin Alice would have to be able to recognize them if they encountered them. When presented with Evans’s Wagera Indian case (Evans, 1973), for instance, they might not know what to say, or they might say the same thing, both of them calling it a case of ‘reference’. In the latter case, of course, they would not both be right. Later, we discuss the nature of the evidence of differences in the extensions. In particular, we resist the suggestion that these differences would have to emerge in competent speakers’ judgments about cases.

these justify Kripkean externalism with respect to proper names and natural kind terms, they provide an equally strong basis for a similar sort of externalism with respect to philosophical terms as well.

Just as names and natural kind terms can be introduced by dubbing, on this view, so can philosophical terms. One introduces ‘water’ by pointing to some water and saying “water,” or uttering something like “Let us call the liquid there ‘water’.” Similarly, one might introduce ‘referring’ by pointing out a certain act of referring and saying “referring,” or uttering something like “Let us call the act he just performed ‘referring’.” The term would then refer to whatever sort of act that act happens to be. And one might introduce ‘knowing’ by pointing out—or singling out by means of a description—a case of knowledge and saying “knowing” or “Let us call the state into which she just entered ‘knowing.’”

Objections to our extension of externalist arguments

First Objection: We all have deeply unconscious knowledge of necessary and sufficient conditions, or at least of necessary conditions, for the application of philosophical terms.

If a Kripkean or Putnamian mixed internalist were to make an exception of philosophical terms, she would have to claim that in order for someone to understand or grasp a philosophical term, she must have in her mind a set of necessary and sufficient conditions, or at least necessary conditions, for falling under that term. She might fall back of some version of what we call the ‘Platonic assumption’, namely, that we all know the “right” necessary and sufficient conditions, or necessary conditions, but this knowledge is buried very deep down in our minds and is therefore hard to make explicit. We think this move is unavailable to the mixed internalist, for two reasons.

First, the mixed internalist who wishes to endorse the Platonic assumption would have to explain why she assumes that we know tacitly what reference, knowledge, personal identity, and so on, consist in, but not, to use some of Putnam’s examples, what pencils, pediatricians, chairs and tigers consist in.¹⁶ If the view is unacceptable in the latter cases, is it not equally unacceptable in the former? The burden of proof here is certainly on someone who thinks the cases are importantly different.

Not only is the burden of proof on our opponents, but one can argue that inference to the best available explanation should lead one away from the Platonic assumption. Hundreds of years of intense philosophical reflection have produced remarkably little agreement. One could of course

¹⁶ See Putnam, 1975.

attempt to account for this phenomenon by means of the Platonic assumption; perhaps our tacit grasp of philosophical concepts is buried more deeply than we have imagined, and further research is required if we are to make it explicit. But why not instead make use of the distinction between deep and superficial features on which Putnam relies in constructing Twin Earth cases? Seen in these terms, philosophers disagree—either about specific cases or about philosophical theories—because while they associate the same superficial features with philosophical terms they differ with regard to the underlying characteristics of what lies in their extension. Their theories attempt to provide necessary and sufficient conditions, or necessary conditions, for the application of such terms when what is available to them—a grasp of the superficial features—does not determine such conditions. The second explanation seems to be the simpler one, particularly for someone who already endorses externalism with respect to some terms.

Second Objection: Some people must know the necessary and sufficient conditions, or necessary conditions, for applying philosophical terms. It might also be objected that our characterization of the information associated with 'reference' that Alice and her doppelganger share is misleading. Surely, a critic may claim, some of us have more in our heads than they do as far as 'reference' is concerned. After all, Putnam himself introduces the doctrine of the division of linguistic labor, according to which those with severely limited information about the extension of a term are able to use the term meaningfully because they defer to those who are more knowledgeable. Maybe philosophers are the experts to whom others defer in order to use philosophical expressions meaningfully.

This objection introduces a version of the Platonic assumption for experts which we believe falls prey to the same objection we leveled at the more democratic version above: lack of consensus in philosophy is best explained in terms of the hypothesis that philosophers lack knowledge, implicit or otherwise, of the deep features of their subject matter. The present objection also misinterprets the role of the division of linguistic labor in Putnam's account. Putnam does *not* require that for a term to be used meaningfully, there must be experts who know what the underlying properties of the items in its extension are. He does state that, in order for 'elm' to be used meaningfully, someone in the linguistic community must associate more information with the word than he, Putnam, does. But he allows that the expression can have a meaningful employment without anyone's knowing what the essential features of elms really are. The same is true of Kripke. He does not require that some speakers know the true nature

of tigers in order for anyone to use 'tiger' to refer to tigers. 'Water' was used meaningfully before anyone knew that it is H₂O. Why should 'reference', say, be any different?

Third Objection: Reference is not a natural kind. Therefore the difference between the extensions of 'reference' on Earth and on Twin Earth cannot be explained in terms of differences in the microstructures of what is in the extensions of the word on the two planets, the way the differences between what is in the extension of 'water' on the two planets can. Since it is the underlying difference in microstructure that explains both the superficial similarities of the liquids as well as their ultimate distinctness, the attempt to extend Putnam's style of argument fails.

For the sake of argument, we can agree that reference is not a natural kind. But we deny that the superficial feature—underlying feature distinction on which the Twin Earth style of argument depends only can be found in connection with natural kinds. Externalist arguments can be given for a variety of non-natural kind terms, ranging from proper names, to what can be labeled 'intentional kind terms'. It is useful to consider this spectrum in order to bridge what for some may be a felt gap between natural kind terms and the philosophical terms with which we are concerned here.¹⁷

One can clearly construct successful Twin Earth cases with names. 'Jack Kennedy' can be used to refer to a person who is not Jack Kennedy, but is in effect a replica of his on Twin Earth, superficially and microstructurally indistinguishable from Jack Kennedy. In such a situation, the referents of 'Jack Kennedy' are obviously different on the two planets. Two speakers, Alice in this world, and Twin Alice on Twin Earth, may be mentally indistinguishable, even though one uses 'Jack Kennedy' to refer to Jack Kennedy, and the other to Kennedy's replica. This shows that Twin Earth arguments require neither that one employ natural kind terms, nor that the referents of the terms differ at the microstructural level.

Next consider a Twin Earth example involving the term 'junk bond'. Junk bonds are conventionally constructed entities. They are the sorts of things that exist in part only because some people have agreed that they exist, and have the properties they have because they have been assigned those properties. Junk bonds do not constitute a natural kind and do not have a microstructure in the same way natural kinds do.

Now, consider a Twin Earth in which the term 'junk bond' is used to refer to financial instruments, call them T bonds, that are superficially similar

17 As we have seen, Putnam mentions pencils, pediatricians, and chairs; Burge (1979) mentions microchips and briskets.

to junk bonds in that they are bought and sold the same way and play the same, in part catastrophic, role in the financial life on Twin Earth. The similarities are such that a moderately interested observer of the financial scene would never notice the difference. There are, however, significant differences. Junk bonds do not necessarily convey an equity in an enterprise, but we stipulate that T bonds differ from junk bonds in that they necessarily include a clause that allows bondholders to convert bonds to shares of the issuer's common stock at a specified conversion value. This makes them significantly different from junk bonds.

Imagine again a speaker on Earth, say Alice, who can use the term 'junk bond' to refer to junk bonds, even though she knows very little about junk bonds, and would be incapable of distinguishing them from T bonds. She resembles the authors of this essay, before we engaged in some economic research for the purposes of constructing this example. All Alice knows about junk bonds is that they have something to do with high finance. Now assume that Twin Alice, on Twin Earth, is psychologically identical to Alice. She associates with 'junk bond' exactly what Alice does. Nonetheless, when Alice uses 'junk bond' she refers to junk bonds, but when Twin Alice uses 'junk bond' she refers to T bonds. Again, this is a Twin Earth case that does not rely on differences in microstructure. Philosophical terms may well belong to the same, no doubt heterogeneous class.

Fourth Objection: Unlike the original, in our extension of Putnam's argument, there are psychological differences between Twin Earth and Earth.

Kripke and Evans tell us that certain complex interrelations between dubbings, the information associated with terms, intentions to use terms the way others do, and so on, determine the referents of certain expressions. Someone might ask about both of these accounts, What makes this a correct description of our linguistic practice? One plausible reply is that whichever account is correct, it is correct as a result of linguistic conventions that developed over a long period of time. These conventions are instituted and upheld by intentional acts; they are produced and maintained by what we do as language users. Hence, the differences between the extensions of 'reference' that we envision in our thought experiment must be accompanied by psychological differences between those who are responsible for instituting the relevant conventions. In this respect, it could be argued, our example is different from Putnam's original Twin Earth case involving 'water', since in the latter case all psychological facts remain the same across the two worlds.

For the sake of argument, let us grant the key claims at issue here. It is plausible to assume that some of the actions and intentions of some of the

members of the linguistic communities in our extension of Putnam's argument differ from those of their doubles. And some of these differences may well reflect corresponding psychological differences.¹⁸ However, the resulting disanalogy between our example and Putnam's is significant in this context only if it implies that in order to use 'reference' meaningfully, speakers need to have in their minds something that determines uniquely identifying conditions, or necessary conditions, for what is in the term's extension. But this implication is not forthcoming. In particular, we see no reason to think that the presumed psychological differences underlying the differing conventions under discussion cannot be spread out in time and across members of the linguistic communities such that (a)–(c) below are all true.

(a) *The psychological differences do not affect Alice's and Twin Alice's internal states at all.* If this is the case, our thought experiment and Putnam's are alike in the crucial respect. In both cases, the focal individual and her doppelgänger are in the same psychological states.¹⁹ Just as S and Twin S can use 'water' without knowing what water is, Alice and Twin Alice can meaningfully use 'reference' even though they do not have in their heads anything that singles out the respective extensions of the term in their worlds.

(b) *The psychological differences are not such that philosophers need have any sort of special access to them.* Like Alice and Twin Alice, philosophers can use 'reference' meaningfully without possessing individuating information about what lies in its extension. In particular, they can do this without being privy to all the conventions that determine its extension.

(c) *There need not be any single individual in a community who has complete knowledge of all the actions and intentions that make it the case that 'reference' has the extension it has in that community.* In other words, there need not be any "reference experts" who know as much about reference—tacitly or otherwise—as chemists know explicitly about water. This is not to say, of course, that there could not be experts about reference. One way to become such an expert would be to engage in a sociolinguistic study of the sort of conventions we have been discussing.

¹⁸ On the other hand, it may be that the supposition that there are conventional—and therefore intentional—differences between the planets in our example is compatible with there being no psychological differences—as long as the psychological states are considered independent of the environment, of course. Externalism, after all, denies that all intentional differences are reflected in psychological ones in this sense. Further argument is required to determine in just what cases such mirroring occurs.

¹⁹ Compare Burge's arthritis example, in which enormous changes in the medical community leave the focal individual unaffected. See Burge, 1979.

It goes beyond the scope of the present paper to *establish* that the relevant difference in conventions is consistent with the truth of (a)–(c). The difficulty of doing this has largely to do with the fact that Kripke and Evans have little or nothing to say about the nature of the conventions that, according to the objection we are considering, make their theories true or false. They do of course provide arguments for their views; but these consist mostly of appeals to intuitions about cases.

Despite the fact that we lack an account of such conventions, we think the burden of proof in the present instance lies with those who think that (a)–(c) are false. Since none of us are consciously aware of just what conventions govern reference, the claim that some or all of us know tacitly what they are stands in need of an argument. For the reasons we gave in reply to the first two objections, we doubt that a compelling argument to this effect can be formulated.

Fifth Objection: Doesn't the fact that we do not now have an account of reference that is as successful as our current chemical understanding of water constitute a glaring disanalogy between Putnam's original Twin Earth thought experiment and our attempt to construct a similar experiment regarding 'reference'?

It is not clear that Putnam's argument turns on the existence of precisely specifiable methods of determining the underlying composition of what is in the extensions of 'water'; therefore, it is not clear that extending his argumentative strategy to philosophical terms requires us to specify such methods. We all know that methods of chemical analysis exist, and we know the correct chemical analysis of water, but does the argument turn on this knowledge? Why would it not suffice for the Twin Earth thought experiments to work merely that we *stipulate* in understandable ways that the extensions of the terms at issue differ? A science of sociolinguistics as successful as modern chemistry may not now exist, but it strikes us as excessively verificationist to demand more than we have provided in the way of a suggestion about how one might find out what the hidden, intentional structure of reference is.

Sixth Objection: How Can There Not Be Reference on Twin Earth?

One might argue that, as we have set up the thought experiment, it is hard to see how there could not be reference on Twin Earth, a situation that makes our case different from Putnam's. The problem arises because in order to describe the thought experiment, we say things like "Suppose that the extension of 'reference' on Twin Earth is correctly described by Evan's theory of reference." In saying this, we are using 'extension', which is linked

semantically to 'reference', the extension of which in this world, we assume for the purposes of our thought experiment, Kripke has described correctly. To put the point loosely, if Evan's theory is a correct account of the *extension* of 'reference' on Twin Earth, how can there be no reference there?

We do not need to use the word 'extension' in the way this objection assumes we do. We can use it instead as a general term that covers different kinds of language-world connections. Our strategy is a familiar one. Someone who is convinced that descriptions and names are importantly different linguistic devices, because they "connect to the world" in fundamentally different ways, can use the expression 'denotation' as a cover term for what they both do.²⁰ 'Denote' is then used as a cover term for both relations. We mean "extension" to be understood in this general sort of way.

General Comment about the Third, Fourth, Fifth, and Sixth Objections

The third, fourth, fifth, and sixth objections to our extension of the externalist arguments are members of a family of objections all of which attempt to show that there is a problematic disanalogy between the two planets in Putnam's example and the two planets in our example. It is crucial to notice that such objections are relevant only to our attempt to extend Putnam's argument. Since Kripke's argument does not require thought experiments about possible worlds²¹—some of his examples involve counterfactual situations, but he also uses actual cases, such as the Feynman case—such objections are irrelevant to our extension of Kripke's argument. So, a general response to all such objections is that unless they are backed up by an objection to our extension of Kripke's argument, all they *would* show if they *were* successful is that *one* argument for externalism cannot be extended to 'reference'.

Seventh Objection: Why think that all philosophical terms can be externalized?

We conjecture that Putnam's and Kripke's arguments are applicable to all philosophical terms. In order to effect the generalization in the Putnamian case, all that appears to be required is, firstly, a distinction between the superficial characteristics associated with what lies in the extensions of philosophical terms and their deep, underlying properties, and, secondly, some basis for concluding that only the superficial properties are "represented" in our heads, or need be in any of our heads. For reasons given

²⁰ See Russell, 1956.

²¹ Nathan Salmon makes this point in his 1981.

above, the history of philosophy provides us with evidence that the second of these requirements is satisfied. And although we wish to remain neutral here regarding the ultimate acceptability of the essentialism at issue in the first requirement, we take it that mixed internalists will find it congenial. Their commitment to it is all that is necessary for our argument to succeed. In the Kripkean case, all that is necessary is that it be possible for one to refer to something without possessing individuating information about it. Despite the force of these general considerations, we think a valuable research program opens up at this juncture, devoted to a case by case evaluation of externalist arguments directed at particular philosophical expressions.

Global Internalism

We conclude that someone who accepts Kripke's or Putnam's arguments for externalism should treat at least some philosophical terms in an externalist manner as well. Mixed internalism is an unstable position, and therefore does not provide us with a construal of the concept model that is any more capable of justifying the intuition method than is the externalism about philosophical expressions we examined earlier.

The remaining alternative is thoroughgoing internalism about all terms. Our discussion of this option will be brief, in part because the externalist attack on global internalism is widely regarded as successful, and in part because a detailed assessment of this critique would take us too far afield. It is worth noting, however, that someone convinced of global internalism is in an awkward position when defending the intuition method by means of the concept model. The intuitions about Kripkean and Putnamian thought experiments on which externalism relies are both prevalent and strong. A thoroughgoing internalist has to succeed in explaining why these intuitions are defective, and must therefore revise the standard version of the concept model, according to which such intuitions are generally reliable.

There is also an important motivation for letting the justification for the intuition method remain neutral between internalism and externalism. If you have to rely on the intuition method to decide between internalism and externalism—and all arguments we know of that purport to accomplish this do rely on it—you would want an independent justification for that method to avoid the vicious circle that would result if you employed the method to justify your favored position and then justified the method by appealing to your favored position.²²

²² While they do rely on the intuition method, since externalists cannot use the concept model as a basis for it, finding such an alternative is particularly important for them.

Part III: Alternative Justifications of the Intuition Method

The first two parts of this paper can be thought of as illustrating a form of argument that can function to challenge any attempt to justify particular applications of the intuition method by means of the concept model. The argument forces a proponent of the concept model either to defend global internalism, or to establish that the class of philosophical terms with which she is concerned is impervious to externalist arguments. Although we have not considered the global internalist option in detail, for the reasons given above we believe it is problematic. And we take it that our arguments present a serious challenge to mixed internalism.

There are of course other approaches to understanding the use of intuitions in philosophy. We lack the space to address them in detail, but, anticipating further discussion, it is worth mentioning and commenting on some of them here. We consider two alternatives. The first takes intuitions to be based at least in part on beliefs with empirical content. The second employs Rawls's pragmatic construal of reflective equilibrium.

Intuitions in philosophy as guided by beliefs with empirical content

One plausible way to construe the view that intuitions in philosophy are based on beliefs with empirical content is as follows:

Intuitions about cases concerning reference, personal identity, knowledge, and so on, are guided by our beliefs about reference, personal identity, knowledge, and so on.²³ Many of these are specific *beliefs*, like the beliefs that Monica used 'Bill' to refer to Clinton, that Clinton is the same person he was before he became president, and that Clinton knows that he is in trouble. Some of the beliefs that guide our intuitions, however, may be general and theoretical. When we are presented with a philosophical thought experiment, these beliefs guide our intuitions by means of largely unconscious deductive and inductive inferences, analogical reasoning, and similarity judgments. Since it is not assumed that the beliefs that guide the intuitions are acquired *a priori*, this does not make the appeal to intuitions a form of *a priori* reasoning.

We will briefly indicate two problems we see with this attempt to defend the intuition method.

First, if you plan to base your judgments about a thought experiment that is crucial to some philosophical theory on the beliefs you already have about the relevant subject matter, you should make sure those prior beliefs are true. There is reason to think they are not. As we pointed out earlier, philosophers have a number of false beliefs, since we disagree and our views cannot be jointly true. If such false beliefs affect intuitions about new cases (and why shouldn't

²³ Other beliefs may of course be implicated as well. Holistic considerations might well apply here.

they if intuitions are guided by beliefs?), these intuitions are not trustworthy.

Even if we know that the beliefs on which these intuitions are allegedly based are true, it remains unclear whether the intuitions themselves will therefore also be true. The current proposal has it that the processes by means of which beliefs yield intuitions include deductive and inductive inference and analogical reasoning. This hypothesis leaves open the question whether the reasoning that actually produces intuitions is valid.²⁴ And the hypothesis itself is of course highly speculative; we still know relatively little about the psychological mechanisms that underlie intuitions.

In sum, the claim that intuitions are guided by beliefs, many of which are empirical, has little or no normative force by itself. It identifies the source of these intuitions, but does not suffice to legitimize their use as data for evaluating philosophical theories. The further claim that these intuitions are reliable is based on assumptions, both about the source beliefs and the psychological mechanisms that generate the intuitions, that we have as yet insufficient reason to accept.

Reflective Equilibrium

There is of a tradition of understanding reflective equilibrium as a method of conceptual analysis. Our arguments against the concept model apply to this construal. But Rawls's approach to reflective equilibrium²⁵ is different. He uses the method to construct a new concept of justice, rather than to analyze a previously existing one. Rawls aims at a state of affairs in which the new concept fits only some of our previous intuitions about justice. The rest are to be jettisoned, not because they are inaccurate, but for practical reasons. In short, Rawls's aim is not to say something *true* about justice and fairness. The point of constructing the new concept is practical; we *need* a better concept of justice, one that can command the allegiance of a diverse constituency, so that we can reach a consensus about how to improve our society.²⁶

Can this approach serve as a general foundation for appeals to intuitions in philosophy? We think not, for two reasons! In the first place, adopting Rawls's conception as a basis for appeals to intuitions in areas in philosophy such as the philosophy of language, epistemology, and the philosophy of mind would require that we give up the idea that theories in these fields are true, in favor of the view that they only serve a practical purpose. We do not believe that many philosophers who work in these areas would be willing to

think of their work in constructive terms, as devoted to devising new concepts constructed from widely shared intuitions. Philosophers in these fields generally believe that their theories are *true*. This leads us to our second reservation. One can plausibly argue that there is a pressing practical need for a new concept of justice; but it is simply not evident that there is a *practical* need for newly constructed concepts of reference, knowledge, personal identity, and the like.

Conclusion

A number of avenues for further research open up at this juncture. One leads to a further investigation of the scope of externalist arguments. Is our conjecture that all philosophically important terms are subject to externalist analyses correct? Another line of inquiry for which our arguments provide a motivation is that of studying the actual psychological mechanisms that give rise to intuitions. What are these mechanisms, and how reliable are they? In particular, if intuitions are informed by beliefs, how exactly does this occur? Which beliefs guide intuitions in the philosophically interesting cases, and are these beliefs true? A related research program is that of attempting to devise other, more promising ways to justify appeals to intuitions in philosophy.

Herman Cappelen
Vassar College

Department of Philosophy
Poughkeepsie, NY 12604
hecappelen@vassar.edu

Douglas G. Winblad
Vassar College

Department of Philosophy
Poughkeepsie, NY 12604
dwinblad@vassar.edu

References

- Bealer, George, "A Priori Knowledge and the Scope of Philosophy," in *Philosophical Studies*, Vol. 81 (1996).
- Burge, Tyler, "Individualism and the Mental," in *Midwest Studies in Philosophy*, Vol. IV, *Studies in Metaphysics*, 1979.
- Evans, Gareth, "A Causal Theory of Names," in the *Aristotelian Society Supplementary Volume XLVII* (1973).
- Gettier, Edmund, "Is Justified True Belief Knowledge?", *Analysis*, 23 (1963).
- Kahneman, Daniel and Tversky, Amos, "On the Psychology of Prediction," *Psychological Review*, 80, 1973.
- Kripke, Saul, *Naming and Necessity* (Cambridge, Mass.: Harvard University Press, 1980).
- Kripke, Saul, "A Problem in the Theory of Reference," in Cauchy, Venant, ed., *Philosophy and Culture*, (1986).

24 See Kahneman and Tversky, 1973 and Tversky and Kahneman, 1974, for evidence that judgements about cases based on statistical reasoning can be systematically mistaken.

25 Rawls, 1971.

26 Rawls, 1985.

- Mackie, J. L., *Ethics: Inventing Right and Wrong* (Hammondsworth: Penguin Books, 1977).
- Nozick, Robert, *Philosophical Explanations* (Cambridge, Mass., Harvard University Press, 1981).
- Perry, John, "Williams on the Self and Its Future," *Journal of Philosophy* 73, no. 13 (1976).
- Putnam, Hilary, "The Meaning of 'Meaning,'" in *Mind, Language and Reality, Philosophical Papers, Vol. 2* (Cambridge: Cambridge University Press, 1975).
- Rawls, John, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971).
- Rawls, John, "Justice as Fairness: Political not Metaphysical" in *Philosophy and Public Affairs*, 1985, Vol. 14.
- Salmon, Nathan, *Reference and Essence* (Princeton, Princeton University Press, 1981).
- Sayre-McCord, Geoff, "'Good' on Twin Earth," in Enrique Vilanueva, ed., *Truth, Philosophical Issues*, 8, 1997.
- Tversky, Amos, and Kahneman, Daniel, "Judgments under Uncertainty: Heuristics and Biases," *Science*, 185, 1974.
- Williams, Bernard, *Problems of the Self* (Cambridge, Cambridge University Press, 1973).

Facta Philosophica 1, 1999: 217-226
© Peter Lang, Switzerland

Die strukturalistische Modellierung der empirischen Wissenschaft

C. Ulises Moulines

Wissenschaftstheorie stellt eine der Disziplinen dar, die sich mit dem Phänomen „Wissenschaft“ beschäftigen. Ähnlich wie Kunst, Religion, Recht, gehört dieses Phänomen zu den kulturellen Erzeugnissen der Menschheit; daraus folgt, daß Wissenschaftstheorie einen Teil der gegenwärtigen Kulturwissenschaften darstellt. Allerdings ist sie eine etwas sonderbare Kulturwissenschaft, da sie eng mit Fragestellungen traditioneller Bereiche der Philosophie (insbesondere der Logik, Sprachphilosophie, Erkenntnistheorie und Ontologie) verknüpft ist.

Genau wie in anderen Zweigen der Kulturwissenschaften kann die Wissenschaftstheorie das Objekt ihrer Untersuchung von zwei grundlegend verschiedenen Gesichtspunkten her ableiten: von einem synchronischen oder einem diachronischen. Diese methodologische Unterscheidung, die auf die Ausführungen von Ferdinand de Saussure über die Verfahrensweisen der Linguistik¹ zurückgeht, ist bei jeglicher Art Forschung äußerst nützlich, nicht nur, um festzustellen, um welches Thema es sich dabei handelt, sondern auch, wie dieses Thema darin behandelt wird. Wenn man die Unterscheidung zwischen einer synchronischen und einer diachronischen Perspektive systematisch aufrechterhält, erspart man sich viele unnötige und unergiebigere Querelen. Viele der vermeintlich fundamentalen Konflikte in der Wissenschaftstheorie seit der sog. „historizistischen Wende“ in den sechziger Jahren gehen ursächlich nur auf eine Verwechslung beider Perspektiven zurück oder auf den monopolistischen Anspruch der einen zum Nachteil der anderen. Auch zu seiner Zeit hatte Saussure große Schwierigkeiten, seine historizistischen Kollegen davon zu überzeugen, daß die synchronische Linguistik eine Existenzberechtigung hat. Heute zweifelt bereits niemand mehr daran. Das gleiche kann man auch von der Wissenschaftstheorie behaupten, die in dieser Beziehung (und in vielen anderen) viele methodologische und begriffliche Parallelen mit der theoretischen

1 F. de Saussure: *Cours de linguistique générale*. Paris, 1916.