16 The Thought Experiments: Reply to Donnellan

Keith Donnellan's essay raises important questions about the nature of the Twin Earth thought experiments. The relations between Putnam's thought experiment in "The Meaning of 'Meaning'" and mine in "Individualism and the Mental" are complex. The obvious similarity is the use of the Twin Earth methodology, for which I am indebted to Putnam. Putnam focused on language, not mind. He did not initially argue for anti-individualism about mental states at all. In fact, he assumed that the psychological states of the twins are the same. I criticized this construal of his thought experiments. I took his thought experiments regarding natural kinds, properly interpreted, to provide a different type of support for the most general anti-individualistic conclusion of my arthritis thought experiment. The most general conclusion is that the nature of many mental states is such that being in those mental states necessarily requires bearing certain relations to one's environment. Putnam and I are now in agreement on these matters (Burge 1982; Putnam 1996).

There are other similarities between my arthritis and water thought experiments. Both depend on the fact that we are not omniscient and the fact that there is possible slack between what we know descriptively about the referents, or correct applications, of our concepts and what their referents are. Since the sort of incomplete linguistic understanding that enters into my thought experiments is equally an incomplete knowledge of the way "the world" is—or what is the true character of what is objectively indicated by our thought and language—there is a sense in which both thought experiments depend on ways that "the world" determines our thought.

I

I think that Donnellan is incorrect in suggesting that I have conflated the two thought experiments, and in suggesting that I have argued that mine is a variant of Putnam's. I shall try to explain this as matters proceed.

Donnellan claims four differences between the thought experiments. The first is that the relevant environment in my thought experiments is social or linguistic, whereas the relevant environment in Putnam's thought experiment is the nature of the physical world. I have, of course, highlighted this point. But I reemphasize that the understanding of language is not sharply separable from knowledge of the world. Not understanding that 'arthritis' applies ("by definition") only to diseases in joints is equally not knowing that arthritis occurs only in joints. I believe that the thought experiments that show that perception and ordinary thought about the physical world are anti-individualistically individuated are in their pervasive application and ontogenetic primacy more fundamental than the social anti-individualistic thought experiments.

Moreover, I think that it is a mistake to hold, as Donnellan appears to, that sofas will have no place in a thought experiment where the relevantly different environments are nonlinguistic. In Burge (1986b), I showed that thought experiments can be constructed wherever there is scope for radical theory about the true character of a type of entity. There is such scope even for sofas, as I specifically argue in that paper.

I agree with Donnellan's account of a second difference between the thought experiments—his account of the difference in the ways experts enter into the two thought experiments. I want, however, to enter a qualification. It is not essential to the "social" thought experiments that the individual rely on "experts" or others with more competence for fixing the linguistic intentional meaning of his terms. One can develop the case in such a way that the patient does not rely on the doctor's meaning. The patient and the doctor can have different meanings for the term 'arthritis', in their respective idiolects, in the actual situation. I think that it is natural and correct in many instances to take them to have the same meaning. But there can be such differences of idiolect, without those differences' undermining the main anti-individualistic conclusion of the thought experiment (Burge 1989).

What is crucial is that the referential application of the patient's term be dependent on the social environment for its connection to the world. The patient's term and the patient's concept can attach to arthritis, even if he and the doctor conceive arthritis differently and mean different things by 'arthritis'. What makes the thought experiment work is that the patient has mistaken beliefs about the disease arthritis, which the doctor's superior expertise reveals to him. In the twin situation, the patient's concept attaches to a quite different syndrome of diseases, as does the doctor's. Given the nonindexical nature of the respective patients' intentional contents or concepts, those contents' having different referents individuates them as different contents. This difference entails that the patients in the different environments are in different mental states. The difference depends on differences in their social environments even though the patients in the two cases do not share concepts with the experts in their environments. They share only referents with those experts. Or at any rate their referents are sufficiently affected by their dependence on others to ensure that the respective patients have different referents from one another.

I am more doubtful about Donnellan's account of a third difference. I do not see any reason why my thought experiments are not parallel to Putnam's in suggesting a "semantical rule," though a different one, that describes the social thought experiments. Something like: "Under such and such conditions, the referential application (or the sense) of the patient's term 'arthritis' is the same as the referential application (sense) of the term as used by others on whom the patient depends in such and such a way." The referential application of the term as used by others will depend on the term and the others. I will return to the bearing of such semantical rules on anti-individualism.

Donnellan's fourth point concerns an indexicality that he sees as central to Putnam's thought experiment but absent in mine. There are several things that Donnellan says on this topic that I believe to be mistaken, or at best misleading. Donnellan agrees in locating indexicality not in the terms themselves. He places it in a semantical rule governing them. He admits, however, that the rule itself does not have to involve indexicals. It could be formulated: "Water is whatever has the important underlying physical characteristics of the stuff users of the language, which itself could be specified nonindexically, call 'water'."[1] We could replace 'users of the language' with 'we', as Donnellan proposes; or we could use a definite description. There are other nonindexical possibilities.

Why does Donnellan feature the indexical formulations? He writes, "We want [them] in part at least in order to ensure that in saying that the psychological states of counterparts on the two planets can be the same. . . ." He is not fully explicit why "we" want such a thing. Sometimes it appears that it is because Donnellan is trying to be faithful to Putnam's original interpretation of the thought experiments. On that original interpretation, the twins were in the same psychological states, but differed in the meanings of their linguistic terms. But I have shown this interpretation to be mistaken. It is now rejected by Putnam as well. The relevant psychological states of the twins are *not* the same.

Sometimes it appears that Donnellan writes what I have just quoted because he himself wants to hold that Putnam's thought experiments do not support anti-individualism. For if the psychological (mental?) states are the same between actual and twin situations, the thought experiment does not support the view that the nature of the individual's mental and psychological states is dependent on relations to the physical environment.[2] Donnellan claims that we can imagine the twin's "having the same history and (in an important sense) psychological states." But he never gives any argument for this contentious point. I believe that the twins cannot have *all* the same psychological states, unless 'psychological' is given a narrow meaning that neither common sense nor empirical science actually uses. I have argued this in various publications, and Putnam has argued similarly.

Suppose that it could be shown empirically—what I doubt—that all relevant individuals have such a rule in their minds or heads. The implausibility of using this line to try to save individualism is highlighted by the fact that perceptual categories, in animals and young children, are subject to relevantly similar thought experiments (Burge 1986a,c). Such thought experiments do not depend on an *underlying* nature but merely on an objective, mind-independent perceptual kind. But perceivers do not carry around in their minds an analogue of the indexical semantical rule governing the application of their perceptual categories.

I find such a supposition about an internal semantical rule doubtful even for the case of natural kind concepts. I think it empirically doubtful that young children, or indeed many adults, who have a distinctive use for natural kind words, and thus a sense for the

appearance–reality distinction, have metaconcepts for words and for calling something by a word. I see no reason why they *must* have such metaresources in order to use natural kind terms in the way we do. Spelling out the rule in detail—specifying what stuff it is that is being called 'water'—requires a conceptual-descriptional knowledge that I think most users of the term or concept simply do not have. The information is often stored perceptually, not conceptually. (For more on this, see my reply to Peacocke in this volume.) I believe at the very least that individuals *need not have* conceptualized their usage in order for that usage to be correctly described by such rules.

   Let us waive all these doubts, however. Suppose that the rule is formulated indexically and that the indexical formulation is common to all relevant twins. Then it still does not follow that the individuals are in the same psychological states. For the words and concepts that the indexically formulated rules govern are *not* indexical. The semantical rule is not synonymous with or conceptually identical with the words or concepts being semantically explicated. The sentences and thoughts using these words and concepts get their intentional content, not just their referents, from interaction with the actual environment that they are in. Since the environments are relevantly different, the intentional contents are different. The intentional contents are distinct because of the difference in the environments that the twins are related to.

   The original point I made in "Other Bodies" that the relevant concepts and words are not indexical is not merely a quick, shallow point. It goes to the heart of the difficulty for defending individualism along these lines. Donnellan gives no grounds for believing that the point applies only to language and not to mind.

II

Donnellan's widget example is nicely posed and contains more than immediately meets the eye. He imagines that we introduce the word 'widget' as applying to any physical object that has all three physical properties that one finds by repeating the following procedure three times: One opens the Shorter Oxford English Dictionary at random and looks (presumably scanning through entries alphabetically) for the first word that designates a property a physical object can have.

   I think that Donnellan is clearly right to reason that the relevant semantic features, or at least semantic rules like those he gives, of terms explained in this way are not sufficient for a term to be a natural kind term. For one thing, there must *be* a unified underlying nature—something that is not guaranteed by the widget rule, or by any apriori rule for ordinary macroterms like 'water'. For another, there must be more to understanding and using the term than mastering its syntactic namelike properties combined with a widget-like semantic rule. With natural kind terms like 'water' there is also a background of apply-

ing a term according to relatively unified identified macroproperties. These, not a random procedure, guide one to an underlying nature.

I think that Putnam never did show that "the meaning of a term [like 'water'] cannot be both in the head and determine the extension of the term." The proposed semantical rules for 'water' and 'widget' are not their meanings. 'Water' is not synonymous with Donnellan's semantical rule for it (Burge 1982; Putnam 1996). I do think that the semantical rule for 'widget' is closer to being *its* meaning, although here too the meanings will differ. The term 'widget' functions as a namelike term that is rigid (assuming that its reference is fixed by *one* sequence of three dictionary look-ups) in a way that the descriptive meaning of the semantical rule is not. But understanding the term requires mastering nothing further.

Imagine a twin who carries out the same procedure that you do in establishing an extension for 'widget'. Imagine that the twin randomly arrives at different properties and a different extension. Would the twin think a different thought when using the term 'widget' than you would? I do not think that the answer matters, because of the artificiality of the term and the fact that it is importantly different from the terms discussed in the original thought experiments. Because the procedure is random, probably most would say that at least before an extension is established, the terms in the twin situations express the same thought-components and meaning. After the procedure is carried out, perhaps intuitions will differ. I am inclined to hold that the twins will express different meanings and thought-components (assuming that the different extensions are fixed by one sequence of dictionary look-ups in each situation). But I would not want to rest a theory on the point.

III

I would like to turn now to some comments that Donnellan makes on my social thought experiments. He characterizes them as resting on three premises. The first claims the possibility of incomplete mastery of intentional contents or concepts. The second is that the standard of complete, or greater, mastery resides in the usage of the individual's speech community. The third is that twins have incomplete mastery of different concepts because they are in different communities that set different standards.

I would like to emphasize (as I do in my reply to Loar) that the thought experiments invite judgments about particular cases. They do not rest on general premises—although the judgments may be guided by principles that can be discerned retrospectively by reflecting on the judgments about cases. I think that all of the premises Donnellan cites are probably true. The actual presentation of the thought experiments in "Individualism and the Mental" does accord with Donnellan's premises, as far as I can see,[3] though I definitely

do not argue from those premises (Burge 1979). Similar but different thought experiments dispense with incomplete linguistic understanding and locate the individual's idiosyncrasy in criticism of the communal meaning or communal understanding (Burge 1986b). And as I have indicated, the community need not provide a standard for the individual's understanding or linguistic meaning, except insofar as it provides a link to the referent of the individual's term (Burge 1989).

I agree with Donnellan's association of incomplete understanding with Plato, and more broadly with the rationalist tradition. I think that he is also correct in seeing categorial notions like *justice*—I would add logical and mathematical notions—as raising complications about the relevant standard for complete mastery. I believe that social thought experiments can apply to terms like these, but the issues are complex, because of the possibility of a fundamental or dominant concept or meaning that has a certain priority in rational thinking. I hope to address these issues elsewhere.

In the last section of his essay Donnellan invokes the metalinguistic strategy to defend individualism against the social thought experiments. I agree with his description of the case that involves total failure of understanding of French words on a menu. I do not quite agree with his description of what is involved in the passage of a name. I think that names are closely associated with metalinguistic paraphrases, but are nevertheless different (Burge 1973, pp. 425–439).[4] In cases of passage of a name, one has a cognitive marker that associates the name with a mental file that tracks the communicative source. I do not, however, differ fundamentally with Donnellan about this case either.

On the other hand, Donnellan's appeal to a metalinguistic paraphrase to account fully for the thoughts the patient expresses with the word 'arthritis' is, I think, unacceptable. According to Donnellan, "what he really believes is only that he has in his thigh the condition called 'arthritis' or something like that." In "Individualism and the Mental" I argued in some detail that the case as I arranged it could not be credibly deflated in these metalinguistic terms (Burge 1979, IIIC). The individual can be quite explicitly not focused on words, and explanations of his behavior will not go through beliefs about words.

I have also emphasized the enormous extent of the phenomenon of incomplete understanding, or reliance on others for meaning and reference. A significant portion of our vocabularies involves terms that we have an imperfect grip on. I find it incredible that metalinguistic beliefs—beliefs analogous to those infected by foreign terms that are not understood at all—underlie the huge range of cases in which what I have called incomplete understanding occurs. We are not cognitively so massively fixed on our own languages. Some of our nonindividualistically individuated beliefs are prelinguistic. Our cognition is largely directed toward the world.

The differences between the ordinary cases I discuss and cases in which one uses a foreign word that one does not understand are palpable. Donnellan provides no answer to

the detailed discussion in my (1979) of the metalinguistic move. So I will not try to take the matter further here. (For more, see my replies to Owens, Peacocke, and Loar.)

I think that Donnellan is right in holding that Putnam's thought experiments, properly understood, help provide new insight into the way natural kind terms are understood: A complete grasp of the concept is still beholden to the empirical nature of the empirical world for determining what is grasped. My social thought experiments bring out how the nature of our thoughts depends on the ways our relation to the physical world is shaped and mediated by our relations to other people and their linguistic and conceptual relations to the same world. But anti-individualism is illustrated in a wide range of cases that are not easily assimilable either to Putnam's natural kind example or to my example of incomplete understanding of terms whose meaning is set by social norms. The slack between an individual's perspectives, physical processes, and individual behavior, on one hand, and the objective environment beyond the individual, on the other, takes many forms. The individual's intentional content or perspective cannot be individuated apart from the objective world that it interacts with and purports to characterize.

Notes

1. The semantical rule needs a fall-back position if there are no unified, underlying physical characteristics. For we cannot know apriori whether a term is a natural kind term. But I omit to enter this into the rule, for the sake of simplicity.

2. I do not see that Donnellan is suggesting a distinction between psychological and mental states. I believe that such a distinction would be pointless and unattractive.

3. A minor exception is that in my examples, it is not obvious that the twin, as opposed to the patient in the original "actual" situation, has an incomplete mastery of his term.

4. I do not regard the predicative aspect of the name 'Aristotle' as synonymous with 'is called "Aristotle"', but I think that the latter explains the core semantical conditions for applying the predicative aspect of the former.

References

Burge, Tyler. 1973. Reference and Proper Names. *Journal of Philosophy* 70: 425–439.

———. 1979. Individualism and the Mental. *Midwest Studies in Philosophy* 4: 73–121.

———. 1982. Other Bodies. In *Thought and Object*, A. Woodfield (ed.). Oxford: Oxford University Press.

———. 1986a. Individualism and Psychology. *Philosophical Review* 95: 3–45.

———. 1986b. Intellectual Norms and Foundations of Mind. *Journal of Philosophy* 83(12): 697–720.

———. 1986c. Cartesian Error and the Objectivity of Perception. In *Subject, Thought, and Context*, McDowell and Pettit (eds.). New York: Oxford University Press.

———. 1989. Wherein Is Language Social? In *Reflections on Chomsky*, A. George (ed.). Oxford: Blackwell.

Putnam, Hilary. 1996. Introduction. In *The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's "The Meaning of 'Meaning',"* Andrew Pessin and Sanford Goldberg (eds.). New York: M. E. Sharpe.