



Burge's New Thought Experiment: Back to the Drawing Room

Kent Bach

The Journal of Philosophy, Volume 85, Issue 2 (Feb., 1988), 88-97.

Stable URL:

<http://links.jstor.org/sici?sici=0022-362X%28198802%2985%3A2%3C88%3ABNTEBT%3E2.0.CO%3B2-M>

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The Journal of Philosophy is published by Journal of Philosophy, Inc.. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/jphil.html>.

The Journal of Philosophy

©1988 Journal of Philosophy, Inc.

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2002 JSTOR

COMMENTS AND CRITICISM

BURGE'S NEW THOUGHT EXPERIMENT:
BACK TO THE DRAWING ROOM

WHEREAS Hilary Putnam has contended that "meanings ain't in the head," at least not entirely, Tyler Burge has gone even further, making the same claim about mental contents. In previous papers,¹ Burge offered several arguments, based on twin-earth thought experiments, against individualistic individuation of types of mental states. Now he has devised a new argument.² Assuming that kinds of states "are individuated by reference to intentional notions," he has given a new reason for claiming that "there are certain relations between an individual and the environment that are necessary to [his] having certain intentional notions" (709). This time Burge relies not on adaptations to intentional contents of Putnam's twin-earth fantasies concerning meaning, but on an original thought experiment. It features "not incomplete [linguistic] understanding or ignorance of specialized knowledge, but nonstandard theory" on the part of the relevant protagonists (709). Having elsewhere countered Burge's earlier arguments,³ here I will address only the new one. I will suggest that it depends, as do the intuitions that motivate it, on reading too much into the language we use to attribute propositional attitudes. In particular, I will challenge Burge's assumption that when we use a term in the 'that'-clause of an attitude attribution literally and correctly, we must be ascribing the notion expressed by the term to the content of the attitude.

¹ "Individualism and the Mental," *Midwest Studies in Philosophy*, iv (1979): 73-121, and "Other Bodies," in Andrew Woodfield, ed., *Thought and Object* (New York: Oxford, 1982).

² "Intellectual Norms and Foundations of Mind," this JOURNAL, LXXXIII, 12 (December 1986): 697-720. Unless otherwise indicated, all references to Burge are to this article.

³ In *Thought and Reference* (New York: Oxford, 1988), ch. 13, I argue that Burge's earlier arguments depend on questionable intuitions (see fn 16 below) or on questionable descriptions of his examples. For instance, in "Individualism and the Mental," *op. cit.*, Burge describes his well-known "tharthritis" example as involving "incomplete understanding of concepts" or "conceptual errors," even though it is more aptly described as involving incomplete understanding of words or linguistic errors. Indeed, sometimes he seems to equate the two, as when he says that "the error is linguistic or conceptual" (*ibid.*, p. 82), though surely these are distinct. Burge offers no positive account of what it is to think with a concept that one incompletely understands. As for me, it is clear how one can use a word one incompletely understands, but I have no idea what it is to think with a concept that one incompletely understands, for I have no idea what it is to understand a concept over and above possessing it.

The new thought experiment is presented against the background of an original challenge which Burge poses to the dogma that necessary truths cannot be doubted. The truths he has in mind are those which “intuitively give the meaning of an empirically applicable term,” such as

- (1) Sofas are pieces of furniture [of such and such construction] made or meant for sitting.

He also challenges the related dogma that understanding a statement like (1) entails accepting it. These dogmas are supported by the following reasoning, which Burge tries to undermine: ‘sofa’ means *piece of furniture . . . made for sitting*; if one understands both phrases, to doubt (1) would be to doubt that sofas are sofas; therefore, one cannot doubt (1) while understanding it—recognizing its meaning entails recognizing its truth. Moreover, so the reasoning goes, “anyone who *appears* to be doubting that sofas are pieces of furniture . . . made for sitting cannot really be doing so” (701). Presumably the person is doubting something else instead, though Burge does not make this corollary of the reasoning explicit.⁴

To undermine this reasoning, Burge presents a provocative conception of individual and collective linguistic activity, according to which “synonymies are grounded in practice: the most competent speakers would use the two relevant expressions interchangeably” (701). He describes “a vast, ragged network of interdependence, established by patterns of deference which lead back to people who would elicit the assent of others” (702). Within this network occurs the dialectical process of giving explications of meaning, or “normative characterizations,” which seeks an equilibrium between formulations and “archetypal applications.” Burge insists that this dialectic concerns both “how correctly to characterize the relevant entities . . . and how to state the meaning of the [relevant] term; . . . examples are ineliminable from our procedures of meaning-giving” (705).⁵

⁴ At one point Burge says that “the key idea is that some necessarily true *thoughts* or *statements* can be doubted” (698; my italics), as if there were no relevant difference between doubting thoughts and doubting statements. The difference, however, is relevant and important; for one can fully understand a thought and yet, because of linguistic ignorance, not fully understand a statement that expresses it. When Burge speaks of “what is doubted,” he often fails to make clear whether he means a thought or a statement expressing that thought.

⁵ “These disputes usually concern . . . [both] how correctly to characterize the relevant entities; whether all chairs have legs or must have legs; . . . [and] how to state the meaning of the term as such; whether according to (by) definition chairs have legs. . . . The second is typically pursued by trying to answer the first” (704/5). This process goes both ways, however: settling the question whether by

Burge's conception of this dialectic has two features essential to his critique of the reasoning outlined above. First, people can have object-level thoughts, e.g., thoughts about sofas, without possessing a correct normative characterization of the term in question, in this case 'sofa'. Second, although such characterizations, as arrived at through this dialectic, provide norms for understanding, their truth does not rest simply on agreement. People can agree and yet be wrong—general acceptance of a normative characterization may be premature. Even though the meaning of a term is constituted by social practice (it is dependent on usage and on people's critical reflection thereon), it need not be in anyone's head.⁶ Yet one's participation in this practice enables one to operate with the notion expressed by the term. In particular, someone with certain strange ideas about sofas can doubt a necessary truth like (1), even though he possesses the notion of sofa.

Such a person, *A*, is one of the protagonists in Burge's new thought experiment. It is designed to show, given the dubitability of meaning-giving normative characterizations, that "social practices are not the only or ultimate nonindividualistic factor in individuating mental states and events" (707). *A*, who has a normal mastery of English, understands the word 'sofa' and has acquired the normal truisms about sofas. However, he comes to doubt these truisms (e.g., he thinks that sofas would collapse under people's weight) and hypothesizes, contrary to (1), that sofas are really works of art or religious artifacts. He even devises an elaborate account of people's misconceptions about sofas. He is prepared to put his hypothesis to the test empirically, proposes some reasonable tests, and, ultimately, once their results come in, even acknowledges that he was mistaken.⁷

definition chairs have legs may lead not only to the conclusion that 'chair' implies having legs, but even to the conclusion that some of the things we have taken to be chairs, have been calling "chairs," are not chairs after all. As we will see, something analogous to this once happened in the case of 'witch', except that *no* so-called "witch" turned out to be one.

⁶ This phenomenon illustrates what R. D. Laing, in *The Politics of Experience* (New York: Ballantine Books: 1967), aptly describes as "conformity to a presence that is everywhere elsewhere" (p. 84), though his examples of such conformity are more pernicious.

⁷ Burge gives plenty of detail to make this story vivid and plausible. If you are not convinced that there could be widespread error about the function of common manufactured items like sofas, adapt to Burge's purposes Putnam's example involving cats and feline robots. Burge remarks that the thought experiment can be made quite general, and he enumerates a wide variety of terms/notions to which it might be adapted (709). Our criticism of the experiment will not exploit the choice of example.

Then there is *B*, *A*'s physically identical counterpart in another world, a world in which the things called "sofas" really are works of art or religious artifacts. Burge calls them "safos," but people in *B*'s world use 'sofa' to refer to them. Now it so happens that *B* hears counterparts of the truisms heard by *A*, but they are, unbeknownst to *B*, jokes, lies, or figurative utterances. *B* takes them literally and accepts them at first, taking safos to be pieces of furniture . . . made for sitting, but eventually develops doubts much like *A*'s. Unlike *A*, however, *B* *mistakenly* thinks that he is challenging common opinion—like everyone else, he correctly thinks that safos are works of art or religious artifacts.

According to Burge, the experiment shows that, even though *A* and *B* are physically similar and in phenomenologically similar situations, their 'sofa'-thoughts do not have the same contents. Whereas "A mistakenly thinks that sofas do not function primarily to be sat upon, *B*'s counterpart thoughts do not involve the notion of sofa and could not correctly be ascribed with 'sofa' in oblique position"; *A*'s thoughts can be so ascribed. In short, "A and *B* have different thoughts" (708).

Do they? Clearly *A* and *B* have thoughts about different things, sofas and safos respectively, but that is not at issue. Nor is Burge claiming that there is a difference between what *A* and *B* think about these things (sofas or safos, as the case may be), i.e., between the predicative contents of these thoughts. Rather, the difference is that, whereas "A has numerous mental events involving the notion of sofa, *B*'s skepticism does not involve thinking of anything as a sofa" (708). So Burge is claiming that *A* thinks of the objects of the thoughts expressible or ascribable with the word form 'sofa' as sofas, whereas *B* thinks of the counterpart objects (his thoughts are expressible or ascribable in *B*'s world with the word form 'sofa') as safos.⁸ We would attribute 'sofa'-thoughts to *A* with ascriptions of the form,

(2) *A* believes that sofas are. . . .

and we could, if we adopted the word 'safo', attribute 'sofa'-thoughts (i.e., thoughts expressible by the word form 'sofa' as used in *B*'s world) to *B* with ascriptions of the form

(3) *B* believes that safos are. . . .

Another way to put Burge's point, though he does not put it this way, would be to say that people in *B*'s world use the form of (2) to

⁸ Burge uses the phrase 'word form' to avoid suggesting that the same *word* 'sofa' occurs in the languages of both worlds.

attribute beliefs to *B*, but would mean *safo* by 'sofa', thereby imputing the notion of *safo* to *B*.

Underlying Burge's interpretation of this thought experiment is the assumption that, if one literally and correctly uses a term in the 'that'-clause of an attitude attribution, one is imputing to the subject the notion expressed by the term and is, further, including it in the content of the attitude being attributed.⁹ We commonly do this all right, but, I claim, not always. When using a form of ascription such as (2), it is our standard practice to use a word like 'sofas' both to refer to sofas, thereby expressing our notion of sofa, and to impute the same notion to the subject in his thoughts about sofas. Our standard practice must be modified, however, when the situation is not standard, as in Burge's thought experiment. For, in that situation, although *A* possesses the notion of sofa, the thought he expresses when he denies (1) by uttering (4),

(4) Sofas are not pieces of furniture . . . made for sitting.

does not involve the notion of sofa. To be sure, he possesses the notion of sofa and that is what the word 'sofa' means, even to him, but in using (4) to say something about sofas, he is not using the notion of sofa to pick out sofas. We may be inclined to use the form (2) to attribute beliefs to *A*, but this inclination can be explained by the fact that *we* are employing the notion of sofa to think of sofas when we use the word 'sofas' to refer to sofas. It is perfectly consistent for us to cancel the implication that *A* thinks of sofas in the way that we do. We can say, "A believes that sofas are . . . , not that he takes them to be sofas."

What makes plausible Burge's contention that ascriptions of the form of (2) literally describe *A*'s beliefs (including the belief that sofas are not pieces of furniture . . . made for sitting) and impute to him the notion of sofa is that *A* does possess the notion of sofa, he does associate this notion with the word 'sofa', and he does use that word to refer to sofas. However, Burge does not consider the possibility that, when *A* uses 'sofa' to refer to sofas, he does not intend to be using the word literally. For, although 'sofa' means *piece of furniture . . . made for sitting*, even to him, he does not take the things to which he is using 'sofa' to refer to be sofas. If he utters (4), he means that so-called "sofas" (what everybody else calls "sofas") are not pieces of furniture . . . made for sitting. He is exploiting the

⁹ More broadly, Burge assumes that our use of 'that'-clauses in attitude attributions inherently embodies a (correct) principle of psychological state individuation. I challenge this assumption in *Thought and Reference*, *op. cit.*, ch. 10, and in "De re Belief and Methodological Solipsism," in Woodfield, ed., *op. cit.*

common, but in his view mistaken, belief that those items are sofas. Indeed, he might well go on to say that he does not take those items to be sofas. Only then would he be using 'sofa' literally.¹⁰

This way of describing the situation with *A* is perfectly natural, as can be seen if we change the example from sofas to witches. Imagine being in New England in the period when deranged women were called witches, where 'witch' meant *woman in league with the Devil* (or something of the sort). Suppose you began to doubt common opinion and came to believe that these women are not in league with the Devil. Your fellow New Englanders would use (5):

(5) *S* believes that witches are not in league with the Devil.

to attribute this belief to you, but we would not—unless we put 'witches' in scare quotes. Your fellow New Englanders take it to be a necessary truth that witches are in league with the Devil, which they take you to reject, but we do not suppose there are any witches at all. Like you, we reject their belief in witches, though not their notion of witch. Yet Burge's position seems to require that (5) attribute a belief literally and correctly and implies that your belief involves the notion of witch. But then what you believe is the negation of a necessary truth,¹¹ in which case it could not be true. Clearly, a better course is to say that (5) attributes to you a belief about so-called "witches," a belief which does not involve the notion of witch, and that you disagree with your fellow New Englanders about what these individuals are. This is not to deny that you possess the notion of witch; indeed, it figures in your belief that there are no witches.

A parallel situation would obtain if *A* were right about sofas, a possibility Burge does not disallow.¹² Then, if people were to de-

¹⁰ There are various ways in which a speaker can exploit linguistic or factual error on the part of the audience [see Kent Bach and Robert M. Harnish, *Linguistic Communication and Speech Acts* (Cambridge, MA: MIT Press, 1979), pp. 85–87]. The situation described here is somewhat analogous to those examples of referential use of definite descriptions where the speaker does not believe the referent to satisfy the description he is using but is exploiting the fact that his audience believes that it does. Suppose you use 'the man drinking a martini' to refer to someone drinking Sprite in a martini glass. You are not using the phrase literally, although you are exploiting its literal meaning and the fact that your audience would take the man to be drinking a martini. Since you intend to be *taken* literally, however, the analogy with *A*'s use of 'sofa' is imperfect—he does not intend to be taken literally. See my *Thought and Reference, op. cit.*, ch. 6, for further discussion of the referential case and of the referential/attributive distinction.

¹¹ Burge might contend that it is not a necessary truth but false that witches are in league with the Devil, and that witches are (really) insane women, but then he would be implying that witches exist.

¹² Burge's conception of linguistic practice explicitly allows for the possibility that we could be systematically mistaken about a certain class of objects, so that the items

scribe *A* as thinking that sofas are not pieces of furniture . . . made for sitting, but works of art or religious artifacts, they would be attributing to him a belief that conflicts with the common conception of sofas, but in this case the common conception would be a mistaken conception of the items called "sofas." Burge needs to show that there is a relevant difference between the witch case, where so-called "witches" are not really witches, and his sofa case, where so-called "sofas" really are sofas. Yet it seems that the only relevant difference concerns the existence of the objects in question (i.e., whether so-called *Ks* really are *Ks*), and the contents of a person's beliefs and the notions involved in them should not depend on the existence or nonexistence of the objects in question. What applies to the skeptic about witches should apply equally to the skeptic about sofas.¹³

Before elaborating my rejoinder to Burge's interpretation of the thought experiment, I should mention that Burge does anticipate certain other rejoinders. They all challenge his claim that *A* literally doubts what he appears to doubt, and suggest that *A* does not really possess the notion of sofa but does possess some other notion instead. One of these rejoinders¹⁴ is similar to mine, except that it refuses to impute to *A* the notion of sofa: "*A* thinks only that *what most people think of as sofas* are works of art or religious artifacts" (711). Burge replies, "*We may assume* that *A* would say that what most people think of as sofas are sofas. . . . What he questions, quite explicitly *we may imagine*, is not whether these things—or what people think of as sofas—are sofas, but whether sofas are what

we call 'Ks' are not *Ks* at all, as we would acknowledge if we knew their true nature. In "Other Bodies," *op. cit.*, Burge allows for such a possibility in an example involving water, though he speaks of it as "epistemic" (p. 114). With sofas, however, it is also metaphysically possible that they not be pieces of furniture . . . made for sitting.

¹³ In this regard consider Burge's contention that, if a doubt like *A*'s "were to prove well founded, the conventional meaning of 'sofa' would be forced to change. But, despite the change, it might remain appropriate, before and after the change, to attribute propositional attitudes involving the notion of sofa. Both before and after, *A* and his opponents would agree that *these are sofas*, . . . having disagreed over whether all and only *sofas* are furnishings of a certain structure made or meant for sitting" (715). Yet in the witch case there was not the analogous disagreement. Moreover, the meaning of the word 'witch' did not change when people became enlightened about so-called "witches."

¹⁴ Burge considers two other objections along the general line just sketched: that *A*, whatever he may say, "has no thoughts literally attributable to him by using 'sofa' in oblique position, but only metalinguistic thoughts about the word 'sofa'" (710), or that he has thoughts involving only a "reduced" notion of sofa (711), e.g., a phenomenal one.

people think they are” (711; my italics). Of course, what Burge claims that we may assume or imagine is just what is at issue. Instead of arguing for it, Burge baldly asserts,

There is simply nothing in our ordinary practices that precludes our taking *A* as literally entertaining the doubts I have ascribed. Whatever impetus there is behind the objections derives not from antecedent practice, but from the feeling that there must be some way of resisting the ascriptions. This feeling derives from habits that stem from background philosophical doctrine (712).

For my part, this feeling derives (even if I am a diehard individualist) from resisting the habits that stem from a background philosophical doctrine taken for granted by Burge. He assumes that, for an ascription of a thought to be literal, each term occurring in the ‘that’-clause must be used to express an element in the content of the thought being ascribed. But that is simply false. The reason, as noted earlier, is that the ascriber can use such a term to express how he, the ascriber, is thinking of the object of the subject’s thought.¹⁵ Thus, one can say of *A*,

(6) *A* believes that sofas are . . . religious artifacts.

and be speaking literally and correctly, even though one is using ‘sofas’ to refer to sofas and not to express how *A* thinks of the objects of the thought being ascribed. It is being denied not that *A* possesses the notion of sofas but only that this notion figures in the content of the thought in question. However, if one said of *A*,

(7) *A* believes that sofas do not exist.

one *would* be using ‘sofas’ to impute the notion of sofa to *A*, but one would *not* be referring to sofas, since one is not attributing to *A* the belief that they, the things everyone else takes to be sofas, do not exist. Notice also that, if *A* uttered (8),

(8) [I believe that] sofas are . . . religious artifacts.

he would not be speaking literally, since even though he is using ‘sofas’ to refer to what in fact are sofas, he does not take them to be sofas and is thinking of them as *what most people think of as sofas* (or as so-called “sofas”). However, if *A* uttered (9),

(9) [I believe that] sofas do not exist.

¹⁵ I develop this and related points about belief sentences in “*De re* Belief and Methodological Solipsism,” *op. cit.*, pp. 129–131, and in *Thought and Reference*, *op. cit.*, ch. 10.

he would be speaking literally. He would not be denying the existence of the things that he takes to be . . . religious artifacts.

Finally, recall that part of Burge's purpose is to show that one can doubt necessary truths that one understands. Thus, he insists that (10):

- (10) *A* believes that sofas are not pieces of furniture . . . made for sitting.

is literally true. I would grant that, but only in the way that (11):

- (11) *A* believes that sofas are not sofas.

is literally true, namely, where the first occurrence of 'sofas' is used to refer to sofas and only the second is used to express part of the content of *A*'s thought. And (11) is literally true in this way, as is evident if one appended 'but . . . religious artifacts' to (11).

Let us conclude our discussion of the thought experiment by considering the situation of *A*'s counterpart *B*, whom we have been neglecting. Recall that *B* is relevantly similar to *A*. In particular, his experience of sofa-like items and his exposure to uses of the word form 'sofa' are the same. But, of course, in *B*'s world these objects are safos and this word form means *safo*. *B* is right about the objects but wrong about popular opinion. Moreover, according to Burge, *B* possesses the notion of safo, not that of sofa, and *B*'s thoughts "do not involve the notion of sofa and could not correctly be ascribed with 'sofa' in oblique position"; *A*'s thoughts can be so ascribed. *A* and *B* have different thoughts, and thus the thought experiment refutes individualism.

Burge's only reason for denying that *B* possesses the notion of sofa seems to be, however, that in *B*'s world the word form 'sofa' means *safo*, as in ascriptions of the form

- (12) *B* believes that sofas are. . . .

Now, of course, people in *B*'s world do use 'sofas' to refer to safos, even in the context of (12), and use it to express the notion of safo. But it does not follow that they would use it, as in (12), to impute the notion of safo to *B*. Assuming they understood his views, they would realize that to him 'sofa' means not *safo* but *piece of furniture . . . made for sitting*. And, although this is what it means to him, if he uttered (13),

- (13) Sofas are not pieces of furniture . . . made for sitting.

he would be using 'sofa' to express not this notion but rather *so-called "sofa"* (as we saw earlier, the same goes for *A*). Moreover,

even though *B* uses 'sofas' to refer to safos, just as everyone else does, and even though he thinks that they are . . . religious artifacts, he does not intend to be using 'sofas' literally (as it is used in his world). For he thinks that 'sofa' means *piece of furniture . . . made for sitting*, and he does not take safos (which everyone in his world uses 'sofa' to refer to) to be pieces of furniture . . . made for sitting. It so happens that he is unwittingly using 'sofa' to refer to what are literally sofas in the sense of 'sofa' operative in his world (safos), but that is accidental; he does not intend to be using 'sofa' in that sense.

On close examination, Burge's new thought experiment loses its momentous import. That is why we did not consider variants of it, although we could have. For, as Peter Unger¹⁶ has shown, tinkering with the details of a thought experiment can dramatically change one's intuitions. We could have asked, for example: What if roughly half the items called "sofas" are pieces of furniture . . . made for sitting and the rest are works of art or religious artifacts? or What if before 1900 the items called "sofas" were works of art or religious artifacts and after 2000 this will again be the case? We did not take up such questions because, rather than vary the details, as Unger might have done, we considered it sufficient to inspect them as given. The intuitions that support Burge's radical position may be initially plausible, but they become undermined once we look carefully at how the relevant terms are used in various contexts. Then we find that nothing radical can be inferred about propositional attitudes from Burge's new experiment with the language used to attribute them.

KENT BACH

San Francisco State University

¹⁶ See Peter Unger, *Philosophical Relativity* (Minneapolis: University of Minnesota Press, 1984), ch. V. Unger observes that, even in the familiar versions of these experiments, there is a tension between "existence beliefs" and "property beliefs." For example, although we all believe that there are cats and that cats are animals, what if the things we have been calling "cats" are actually robots controlled from Mars? Either there are no cats or cats are not animals. My intuitions favor the former option, Putnam's the latter. Unger plays no favorites, thinking that a good case can be made for either option.