Scientific Research

Mental Content Externalism and Social Understanding

Halvor Nordby[1,2]

[1]Faculty of Education and Social Work, Lillehammer University College, Lillehammer, Norway
[2]Department of Health and Society, The University of Oslo, Oslo, Norway
Email: halvor.nordby@medisin.uio.no, halvor.nordby@hil.no

Tyler Burge has in many writings distinguished between 1) mental content externalism based on incorrect understanding and 2) mental content externalism based on partial but not incorrect understanding. Both 1) and 2) have far-reaching implications for analyses of communication and concept possession in various expert-layperson relations, but Burge and his critics have mainly focused on 1). This article first argues that 2) escapes the most influential objection to 1). I then raise an objection against Burge's argument for 2). The objection focuses on Burge's claim that a person with a partial understanding of a term in our community expresses our standard concept because he is willing to defer to our standard understanding, while his "Putnamian" twin in a counterfactual community does not. The problem with Burge's argument for this claim is that he does not consider the possibility that the person in our community and the twin would defer to the same understanding if they were presented with the same alternatives. Drawing from widespread dispositional assumptions about meaning, I argue that Burge must accept that they express the same concept if they would defer to the same understanding. The article closes with an examination on various ways the externalist may attempt to avoid this problem and concludes that none of them succeeds.

Keywords: Content Externalism; Concept Possession; Partial Understanding; Linguistic Meaning; Conceptual Role Semantics

Introduction

Tyler Burge's influential arguments for social externalism about mental content have their basis in thought experiments involving three steps. In the first step Burge claims that a person in our linguistic community has beliefs involving our standard concept despite having an incomplete understanding of the expression he uses to express that concept. In the second step Burge imagines that the person in our community has a "Putnamian", non-intentionally specified identical twin in a counterfactual community where the same expression has a standard meaning that does not correspond to the correct understanding in our community.[1] Except for this difference we are to conceive of the counterfactual community as identical to ours. The third step is the externalistic conclusion: because of the difference between the two linguistic communities, the person in our community and his twin do not associate the same concept with the expression (Burge, 1979: pp. 77-79).

In their discussions on social externalism, Burge and his critics have paid most attention to the first step of the thought experiment, particularly in connection with Burge's well-known "arthritis" case involving a patient who mistakenly thinks that "arthritis" applies not only to inflammations of the joints, but also to inflammations elsewhere that have spread from the joints (Pessin & Goldberg, 1996; Schantz, 2004). Many of Burge's critics have argued that the patient's concept is not *arthritis* but rather an alternative concept with an extension that directly matches the patient's incorrect understanding of "arthritis", i.e. a concept that applies to inflammations that have spread from joints as well as inflammations of the joints (Woodfield, 1982; Loar, 1987; Crane, 1991; Bach, 1994; Nordby, 2004).

This objection centers particularly on incorrect understanding. Burge, however, also focuses on partial but only incorrect understanding, which he explains as follows:

> One need not rely on an underlying misconception in the thought experiment. One may pick a case in which the subject only partially understands an expression. He may apply it firmly and correctly in a range of cases, but be unclear or agnostic about certain of its applications or implications which, in fact, are fully established in common practice (Burge, 1979: p. 82).

In such cases of partial understanding there is no "underlying misconception", and so the idea about an alternative concept that directly matches an incorrect understanding cannot prevail. Burge's analyses of partial understanding are not as developed as his analyses of incorrect understanding, but they deserve attention both because they are invulnerable to the standard objection to social externalism based on incorrect understanding, and because it is reasonable to assume that partial understanding is a more widespread phenomenon than incorrect understanding. In fact, in many areas of discourse it seems to be the rule and not the exception that a partial understanding is involved.[2] Consider, for instance, doctor-patient interaction and the fact that patients normally have an incomplete understand-

---

[1]Burge's idea of a non-intentionally specified identical twin is adopted from Putnam's "The Meaning of Meaning" (1975). Unlike Putnam Burge makes it clear that his thought experiment focuses on mental content. Putnam has in more recent writings made it clear that he thinks content externalism includes mental content (Putnam 1996).
[2]The latter claim is reasonable for the following reason: normally, when we have beliefs about a term such that we do not have a complete understanding then these beliefs are formed on the basis of correct information from competent speakers or other reliable sources like a dictionary. None of the arguments in this article will depend on the plausibility of this empirical claim, but I mention it since it is relevant for the wider significance of the discussion.

ing of medical terminology. Does this mean that they possess alternative concepts, or do they have the same concepts as the doctors they encounter? Externalism based on partial understanding has considerable implications for how we should conceive of communication and concept possession in this and other forms of interaction between professional experts and laypersons.[3]

This article's central argument is that Burge's argument for social externalism based on partial understanding fails to be convincing even though it avoids the most influential objection to social externalism based on incorrect understanding. The problem with Burge's argument is that there seems to be no good reason why deference-willingness to normative meaning should be restricted to the relation between a speaker and his actual linguistic community. For Burge, the reason why the person with the partial understanding in our community possesses our normative concept is that he is willing to defer to the correct understanding in our community. But why not assume that there is an alternative concept that better matches his own, and that he would have, therefore, chosen this alternative understanding if he were confronted with it?

Granted, since a person with a partial understanding of a term is agnostic about its full applications conditions, no alternative concept can directly match his idiosyncratic understanding as in cases of incorrect understanding like the "arthritis" case. But in the light of widespread dispositional assumptions about meaning, Burge seems committed to accepting that if the person in our community would have chosen to defer to an alternative understanding even though it does not directly match his own understanding, then he has an alternative concept. I will argue that the fundamental problem for the externalist is that as long as the person in our community and his twin are defined as identical "from the inside", it is not unreasonable to assume that they would defer to the same meaning if they were presented with the same alternatives. But then, according to dispositional approaches, they also posses the same concept and thus social externalism is false. I will conclude that there seems to be no satisfactory solution to this problem for the social externalist, and that Burge's argument for social externalism based on partial understanding is, therefore, unconvincing.

The "Arthritis" Case and Cases of Partial Understanding

Burge's well-known "arthritis" case is the most commonly discussed illustration of his notion of social externalism. Burge first asks us to imagine a patient who mistakenly thinks that the word "arthritis" applies to an inflammation he has developed in his thigh as well as inflammations in joints. According to Burge, since the patient is willing to defer to the medical explanation of "arthritis" he has our standard concept of arthritis even before his understanding is corrected: "The patient believes that stiffening of the joints is a symptom of arthritis, that certain sorts of aches are characteristic of arthritis ... and so forth. In addition to these unsurprising attitudes he thinks falsely that he had developed arthritis in his thigh" (Burge, 1979: p. 77). In the second step of the thought experiment Burge imagines a twin of the

patient who lives in a counterfactual community where "arthritis" has a standard meaning that corresponds to the patient's misconception. The third step is the externalistic conclusion: "We suppose that in the counterfactual case we cannot correctly ascribe [to the twin] any content clause containing an oblique occurrence of the term 'arthritis'" (Burge, 1979: p. 78).

Why should we accept the first step of this argument? Why not suppose that the patient expresses an alternative concept that matches his incorrect understanding of "arthritis"? This supposition is what many critics of Burge have put forward. Tim Crane has, for instance, argued that the patient's beliefs involve the concept *tharthritis* "which applies to both arthritis and whatever is the disease he has in his thigh" (Crane, 1991: p. 18). The reason, Crane claims, is that until the patient.

> ...is able to correct his belief about the meaning of the word "arthritis", there is no reason as yet to suppose that he can discriminate between arthritis and tharthritis. So his concept—which at the very least must reflect an ability to discriminate—will apply to arthritis and tharthritis alike (Crane, 1991: p. 19).

Kent Bach has also argued in a similar way. He asks which concept the patient possesses, *arthritis* or some concept that "just captures the misconception" (Bach, 1994: p. 268). Bach then claims that the "very inferences [that the patient makes] constitute the evidence for attributing to the patient some other notion than that of arthritis" (Bach, 1994: p. 268).

These and many other objections to Burge's "arthritis" case have had the following form. 1) The patient's concept is the concept that directly matches his incorrect understanding. 2) There is an alternative concept that directly matches the patient's understanding of "arthritis". 3) The patient's concept is, therefore, this alternative concept. Burge attempts to respond to this objection (Burge, 1979: pp. 110-112), but it has been a widespread view that this response is unconvincing (Woodfield, 1982; Loar, 1987; Jacob, 1987; Crane, 1991; Bach, 1994; Pessin & Goldberg 1996; Schantz, 2004).

Burge's critics have often seemed to think that if one has refuted Burge's "arthritis" argument, then one has shown that social externalism is implausible. The plausibility of social externalism has, in fact, been equated with the plausibility of social externalism based on incorrect understanding, like the "arthritis" case (Pessin & Goldberg, 1996, Schantz, 2004). However, this inference is unjustified. The reason is that Burge makes it clear in several places that he thinks that a partial, but not incorrect understanding, is a sufficient foundation for social externalism. As an example, he mentions the following case:

> [A] protagonist is unsure whether his father has mortgages on the car and house, or just on the house. He is a little uncertain about exactly how the loan and the collateral must be arranged for there to be mortgage, and he is not clear about whether one may have mortgages on anything other than houses. He is sure, however, that Uncle Harry paid off his mortgage (Burge, 1979: pp. 82-83).

It is important to note that as Burge understands this and other cases of partial understanding, an objection similar to the objection against the first step of the "arthritis" case cannot succeed. For when a person has a partial understanding there is no basis for holding that the person has an alternative concept that directly corresponds to his misunderstanding. According to Burge's distinction between incorrect and partial understanding,

---

[3]This expert-layperson distinction is first and foremost associated with Putnam's "The meaning of 'meaning'" (1975), and Burge makes it clear that his arguments are in many ways based on assumptions Putnam makes (1979). A rather different way of developing Putnam's distinction can be found in Evans (1982).

if a person has an incorrect understanding then he has an "underlying misconception", and he does not, then, have a partial understanding.

It is not entirely clear whether this way of conceiving of the distinction corresponds to the understanding speakers of English normally have. In an intuitive sense, a competent speaker might think that someone who has an "incorrect" understanding can sometimes be on "the right track", and in this sense also have a "partial" understanding. As an example, one might refer to cases like the classical "Blind men and the Elephant" story, involving a man who caught hold of an elephant's tail and erroneously believed that an "elephant" referred to something in part shaped as a rope. Some might suggest that this man was on the "right track", because part of an elephant (e.g. its tail) is indeed in the shape of a rope.[4]

It would fall outside the present argumentative purposes to discuss in detail how the distinction between partial and incorrect understanding should be defined, and how it corresponds to common sense. The important point here is that Burge's conception is reasonably clear, and that this is the conception I will use here. Thus, Burge regards a partial understanding as an incomplete understanding that does not involve a mistake about the meaning of the term in question. In the following I will adopt this way of distinguishing between partial and incorrect understanding.[5]

Obviously, this terminological distinction does not imply that persons with an incorrect understanding do not think that they have a partial understanding. In Burge's "arthritis" case it is natural to assume that the patient thinks there is much he does not know about the meaning of the term "arthritis". It is just that he has, for some special reason, started to believe that arthritis is a disease that can occur outside joints. Yet, before the patient defers to his doctor's understanding as Burge assumes that he does, the patient does not think that his belief is incorrect. The patient is nevertheless happy to accept that he does not have expert knowledge about arthritis, that he is agnostic about the full application conditions of the term. Similarly, persons with a partial understanding will normally not consider themselves to have an incorrect understanding. If we think that a belief about the meaning of a term is false (and if we are sufficiently rational), we do not form that belief.

Furthermore, persons with a partial understanding may have partial understandings that differ. Cases of partial understanding can be different both in the sense that they can represent different approaches to a complete understanding and in the sense that they can be more or less partial. But all cases of partial understanding are similar in that they are "on the right track". As long as this is so, then no objection that appeals to an alternative concept that directly matches an "underlying misconception" can succeed.

This observation is not dependent on whether or not the objection succeeds in cases of incorrect understanding. The point is that even if the objection succeeds in cases of incorrect understanding, it does not undermine social externalism grounded in partial understanding. This, at any rate, is what Burge would seem to think. His systematical defense of social externalism is restricted to defending the idea that the person in the "arthritis" case does not possess an alternative concept that matches his incorrect understanding (Burge, 1979: pp. 89-103). Burge writes: "I shall have little further to say in defense of the second and third steps of the thought experiment. Both rest on their intuitive plausibility, not on some particular theory" (Burge, 1979: p. 88).

Burge also says "I find that most people unspoiled by philosophical training regard the three steps of the thought experiment as painfully obvious" (Burge, 1979: p. 87). Burge assumes that social externalism is plausible unless there are good reasons for revising our ordinary use of psychological explanations that ascribes concepts that are incorrectly or partially understood. In the end he summarizes his discussion like this: "Of course, usage is not sacred if good reasons for revising it can be given. But none have been" (Burge, 1979: p. 102). The only objection Burge has considered is the objection that appeals to an alternative concept that matches the incorrect understanding in the "arthritis" case; we must, then, assume that Burge thinks that this objection is the only real threat to social externalism. Burge does not think that social externalism based on partial understanding needs an independent, thorough defense.

But is this really so? It is important to bear in mind that social externalism does not merely rely on how well a person's understanding approximates a complete understanding; it also relies heavily on deference-willingness to normative meaning in this relation An important assumption in Burge's "arthritis" case is that the patient accepts the doctor as an authority on the meaning of "arthritis": "[When told about his mistake] the patient is surprised, but relinquishes his view and goes on to ask what might be wrong with his thigh" (Burge, 1979: p. 77). The same point applies in cases of partial understanding. For the externalist, the reason why a person with a partial understanding has our concept in the first step of the thought experiment, is that he has an understanding that approximates a complete understanding sufficiently, and that he is willing to defer to our normative meaning upon confrontation. And just as in the "arthritis" case, the crucial idea is that the person has our concept already, even before he learns more about its meaning.

Similarly in the third step of the thought experiment, the reason why the twin does not possess our concept is that he is member of a community where the standard meaning of the term is different from the normative meaning in our community. The externalist claims that since our understanding is not available for the twin to defer to, and since he merely has a partial understanding compared to ours, it cannot be correct to hold that he has our standard concept.

If we are going to find a convincing objection to social externalism based on partial understanding, we must address some of these assumptions. But appealing to the idea of a concept that directly matches an incorrect application cannot undermine any of them. If this is the only objection to social externalism we can find, then social externalism based on partial understanding remains an eligible alternative with striking implications for how we should think of concept possession, not only limited to philosophy of mind, but in all areas where questions about language mastery and the relation between mind

---

[4]I would like to express my thanks to a referee for this journal, for mentioning this as an example that illustrates, in a very illuminating way, that it is not obvious how the distinction between partial and incorrect understanding should be fine tuned.

[5]Obviously, much more can be said about these concepts, and it is not, perhaps, clear that Burge's distinction corresponds to all of our intuitions about the meaning of the expressions "incorrect understanding" and "partial understanding". However, discussing this would fall outside the argumentative purposes here. The important point is that Burge's distinction is reasonably clear, and that it can be used to distinguish between the two forms of social externalism.

and world are relevant.

In the following I will develop a stronger objection to Burge's argument for social externalism based on partial understanding. The key idea of this objection is that Burge's concept of deference-willingness as a relation between a person and his actual community is too narrow. Consider traditional dispositional accounts of understanding (Kripke, 1982; Cummins, 1989; Guttenplan, 1994; Loewer, 1996). According to dispositional approaches, how a person understands an expression involves more than how he actually applies it. Considerations about how the person is disposed to use it in counterfactual circumstances are also crucial. Let us suppose that a person with a partial understanding in the first step of Burge's thought experiment defers to our understanding. But suppose also that the reason he does this is that he has not been confronted with other conceptions of what the word in question means. How can we be certain that he would not choose to defer to some alternative understanding that he thinks better matches his own? This, I will argue, is a question Burge pays insufficient attention to.

Deference-Willingness and Implicit Conceptions

A distinction Christopher Peacocke has made between possession conditions and attribution conditions of concepts can serve as a starting point for a deeper examination of Burge's assumptions about deference-willingness.

In Peacocke's influential *A Study of Concepts* (1992), possession conditions of concept are not simply conditions for having beliefs involving concepts. Rather, they are conditions not only for having a concept, but also for having a full mastery of it (1992: pp. 27-33). The reason Peacocke makes this qualification of full mastery is that he wants to leave room for what he calls "attribution conditions", i.e. conditions for truly ascribing a concept to someone who does not have a full command of it:

> The attribution conditions for *red*, the conditions under which something of the form "x believes that _red_" is true, are much weaker than the possession condition. The following are jointly sufficient for such an attribution to be true.
> 1) The subject is willing sincerely to assert some sentence of the form "_red_" containing the word "red" (or some translation of it).
> 2) He has some minimal knowledge of the kind of reference it has (e.g. that it is a color word).
> 3) He defers in his use of the word to members of his linguistic community (Peacocke, 1992: p. 29).

For Peacocke, this idea of attribution conditions is precisely based on Burge's arguments for social externalism. Peacocke claims that the idea that a minimal understanding can be sufficient for true concept ascriptions "has been extensively defended by Burge. I take his arguments as compelling" (Peacocke, 1992: p. 29).[6] Peacocke also writes about his above condition "c" that deference is

> … what distinguishes the case we are interested in, partial understanding (and partial misunderstanding) of a word in

a communal language, from the quite different case of an individual's taking over a word from his community and using it in his own individual, different sense (Peacocke, 1992: p. 29).

Burge makes it clear in several places that he thinks deference is important for the same kind of reason. Towards the end of his discussion he sums up the general idea behind step one of social externalism like this:

> Crudely put, whenever the subject has attained a certain competence in large relevant parts of his language and has (implicitly) assumed a certain general commitment or responsibility to the communal conventions governing the language's symbols, the expressions the subject uses take on a certain inertia in determining attributions of mental content to him. In particular, the expressions the subject uses sometimes provide the content of his mental states and events even though he only partially understands, or even misunderstands, some of them. Global coherence and responsibility seem sometimes to over-ride localized incompetence (Burge, 1979: p. 114).

Burge does not say explicitly that he thinks deference is necessary for the reason Peacocke gives, but we may reasonably suppose that considerations along the same lines must underlie Burge's emphasis on deference. Thus, in the third step of the thought experiment, Burge's idea seems to be that the standard understanding in the counterfactual community is not our correct understanding, and so the twin or any other member of that community is not able to defer to our understanding. According to the social externalist, the twin and the other speakers are in the same kind of situation as the person Peacocke described: they are using the word in question in their "own different sense", compared to the normative meaning in our linguistic community.

It is important to bear in mind that for Burge, this condition of deference only applies to persons who do not have a complete understanding. Burge's arguments are consistent with accepting that a person in the counterfactual community has our concept if he has an understanding that corresponds to a complete understanding in our community, but this issue is not something Burge focuses on in his discussion of social externalism. More generally, Burge's aim is not to present an overall theory of what it is to possess a concept; he is not even interested in analyzing what it is to have a complete understanding. In Peacocke's *A Study of Concepts*, on the other hand, the attribution conditions are conceived of as an amendment to the general theory of what it is to have a full mastery of a concept.

In more recent writings Peacocke has developed a more comprehensive theory of how it is possible to have beliefs involving a concept despite not having a full mastery. According to this theory, a person can have a concept even though he merely has an "implicit conception" of its meaning. Furthermore, Peacocke holds that an "implicit conception" can be sufficient for a person to have a concept even though the correct explanation of the concept cannot be found in the person's actual linguistic community. As an example Peacocke mentions the concept of the limit of a series:

> One of the most spectacular illustrations of this is given by the famous case of Leibniz's and Newton's grappling with the notion of the limit of a series, a notion crucial in the explanation of the differential calculus. It would be a huge

---

[6]As said above, Burge's defense of social externalism is generally focused on the "arthritis" case and incorrect understanding. Burge, in fact, presents no argument that exclusively focuses on partial understanding. Many critics of Burge would therefore disagree with Peacocke when he holds that Burge's arguments are compelling.

injustice to Leibniz and Newton to deny that they had the concept of the limit of a series… I would say that each of these great thinkers had an implicit conception, which explained their application of the phrase "limit of…" in making judgments about the limits of particular series of ratios. What they could not do, despite repeated pressing by critics and well-wishers, was to make explicit the content of their implicit conceptions (Peacocke, 1998: p. 49).

Some philosophers have argued that Peacocke has offered a plausible analysis of cases of this kind (Higginbotham, 1998; Toribo, 1998), while others have been more critical (Schiffer, 1998; Rey, 1998). While a detailed discussion of Peacocke's theory of implicit conceptions would fall outside the present argumentative purposes, we may draw attention to the important point that it is possible to consider the idea of deference-willingness to normative meaning as having a wider application beyond those cases where a person defers to speakers in his linguistic community. According to this broader interpretation, someone who would have deferred to the correct understanding, even though it cannot be found in his actual linguistic community, should also be thought of as someone who has our normative concept: "To whom were Leibniz and Newton supposed to defer? There was no one else who understood the notion better" (Peacocke, 1998: p. 50). Peacocke's crucial assumption is that these theorists would have accepted the modern explanation of the concept of the limit of a series if they had been confronted with it and had adequate time to think about its meaning.

As a further if somewhat different example, consider a linguist who studies a historical language that has ceased to be a living language, i.e. one not presently used in any actual community. The linguist's knowledge of the language, we may imagine, draws on archeological material. For many expressions of the language he has formed a competent understanding, but for a certain expression "c" he believes that he merely has a partial understanding. He is confident that the expression "c" applies to a large carnivorous mammal, but he is not sure which one. Furthermore, no one in his linguistic community knows more about the meaning of "c" than he does, and so there are no linguistic experts whom he can turn to. One day, however, he comes across with some other written inscriptions that settle his uncertainty: it becomes clear to him that "c" refers to male lions. Accordingly, the linguist defers to this understanding of "c".

The problem for the social externalist now begins to emerge. Remember that for Burge, the crucial idea is that persons with an incorrect or partial understanding can have our concepts even before they defer to our understanding. They have our concepts as long as that they are deference willing—as long as they would have deferred to our understanding. It is therefore not clear that the social externalist would be justified in rejecting the idea that the linguistic in the above example associates the concept *male lion* with the expression "c" even before he found the further inscriptions. For in this case deference-willingness is also crucial, the only difference is that the norm exists as written inscriptions and not as a public meaning. But what if we were to go one step further and assume that there is not even a hidden written inscription, but merely a hypothetical explanation of meaning to which the linguist would have deferred?

The point is that we can grant the social externalist that if our understanding is the only understanding a person with a partial

understanding in our community would be willing to defer to, then he also has our concept. But this is to beg the real question. The crucial question is whether we may reasonably assume that this is the choice he would make were he confronted with other hypothetical meaning explanations as well.

Deference-Willingness in the Thought Experiment

Does Burge respond to this challenge? It would be unfair to Burge to claim that he does not attempt to argue for the first step of the externalistic thought experiment. But as noted above, Burge's discussion of social externalism is first and foremost devoted to a defense of the idea that a person in our community can possess our standard concept despite having an incorrect understanding. Furthermore, Burge's discussion of deference-willingness focuses on the internal relation between a person and the normative meaning in his actual linguistic community. Burge does not consider as a possible objection to the first step of social externalism the idea that a person with a partial understanding would have deferred to an alternative understanding.

It is not unreasonable to insist that the externalist should address this objection. After all, most philosophical accounts of language mastery and understanding have been dispositional. It has been a traditional view that questions about concept possession should be answered not only on the basis of reflections on how a person actually applies a concept, but also on assumptions about how he would have used it (Kripke, 1982; Cummins, 1989; Loewer, 1997; Fodor, 1998; McManus, 2000; Kusch, 2005). In fact, a striking feature of Burge's discussion of social externalism is that the understanding of the counterfactual community, as he describes it, could be a candidate for the alternative understanding to which the person in our community would have deferred. Consider again the "mortgage" case. As shown above, the person with the partial understanding in our community is conceived to be a person who "is not clear about whether one may have mortgages on anything other than houses" (Burge, 1979: p. 83). The second and third steps of the thought experiment are described like this: "Imagine our man constant in the ways previously indicated and that "mortgage" commonly applied only to mortgages on houses. But imagine banking practices themselves to be the same. Then the subject's uncertainty would plausibly not involve the notion of mortgage" (Burge, 1979: p. 83).

But what if the person in our community were presented with this alternative meaning of "mortgage"? As shown, Burge claims that our ordinary practices of ascribing beliefs support social externalism. Elsewhere he says that he has "...presented the experiment as appealing to ordinary intuition" (Burge, 1979: p. 88). But it does not seem evident that the person in our community would have chosen our understanding, and if he were to choose to defer to the alternative understanding, then it seems false that "ordinary intuition" supports the view that he has our concept.

What if the externalist responded to this challenge by reformulating the understanding of the person in our community so that he would, on the basis of this understanding, recognize our public understanding as the normative meaning? A person with a well-developed partial understanding of this kind, Peacocke would say, has an implicit conception of the correct explanation of his concept. Burge has, in fact, also focused on this kind of incomplete understanding, although not in his discussion of

social externalism and not explicitly as implicit conceptions:

> Suppose I explicate my word "chair" in a way that requires that chairs have legs, and then come to realize that beach chairs, or deck chairs bolted to a wall, or ski-lift chairs, are counter-examples. Or suppose that I learn more about how to discriminate water from other (possible or actual) colorless, tasteless, potable liquids. In such cases, I learn something about chairs or water that I did not know before. In these cases it is simply not true that the reference of my words "chair" and "water" must change. Although it is true that my conception—my explication—changes, it remains possible for me to observe (with univocal use of "chair"): "I used to think chairs had to have legs, but now know that chairs need not have legs". It remains possible for me to have thoughts about water as water, knowing that there might be some other liquids that I could not, by means other than use of my concept *water*, discriminate from it. Thus there is a sense in which the concept, and the translational meaning of the word in the idiolect, remain the same despite the changed discriminating ability, or change in explication (Burge, 1989: pp. 182-183).

Burge's point is that in cases of this kind a person will, on the basis of the underlying understanding he already has, adjust his explicit statements when confronted with uses that do not match these statements. In this sense he already has an implicit conception that causes him to adjust his meaning explanations. It is this conception, and not his explicit statements about the term, that constitutes the normative standard for him.

The idea in the first step of the externalistic thought experiment would then be this: a person has an implicit conception of the correct meaning of a term in our community. Since the person would recognize our understanding as the understanding that best matches his idiosyncratic understanding, he would defer to our understanding even if he were presented with alternative explanations. This means that it is our concept that he possesses.

However, even if we grant the externalist this line of reasoning in the first step of the thought experiment, the same problem arises in the third step. For if we think of the twin in the counterfactual community as having this kind of partial understanding, then we may plausibly assume that he would have deferred to our understanding as well, for now he has a well-developed understanding compared to our normative meaning. So why not suppose that he would have thought that we are experts about the application of his term; that we are "producers" of a word he is a "consumer" of. The problem for the externalist thus reemerges. Either he needs to explain why the significance of deference should be restricted to the relation between a speaker and his actual linguistic community, or he needs to show that the twin would have chosen his own community.

The Fundamental Problem

The fact that it is not obvious that the person in our community would have deferred to our community is merely one aspect of the most fundamental problem for the social externalist. For just as we can ask whether the person in our community would defer to the counterfactual community as Burge describes it, we can also imagine other communities that are more or less similar. The basic problem is thus: the social externalist needs to show that the person in our community and his twin would not have deferred to the same understanding if they were presented with the same choices. But as long as they are defined as identical "from the inside" in the Putnamian sense—that is, as long as they are supposed to have the same idiosyncratic perspective on the world—it is not unreasonable to assume that Putnamian twins would have deferred to the same understanding. As shown, the crucial idea for Burge is the idea of a person who has "assumed a general commitment or responsibility to the communal conventions governing the language's symbols" (Burge; 1979: p. 114). In accordance with this assumption, if the person in our community and the twin would have deferred to the same understanding, then it seems that this understanding is the best candidate for capturing the nature of their concepts.

It is important to bear in mind that this objection does not depend on the idea of an entire alternative twin earth community to which the speakers would defer. Consider again Burge's "mortgage" case. We may imagine a group of speakers in our community who have chosen to understand "mortgage" in a way that is fairly similar to the understanding of the person in the first step of the thought experiment but is nevertheless different from our standard understanding. How can we be sure that the person would not have deferred to these speakers? We could then introduce the same group of speakers in the counterfactual community in the third step of the thought experiment and ask the same question about the twin.

This point is significant, since some might object that it is not plausible to assume that the person with the partial understanding in our community and his twin would have deferred to alternative twin earth communities. But this is to misunderstand the problem. The crucial question is not whether the person in our community and his twin would have deferred to the same twin earth community. Rather, the crucial question is whether there is some alternative understanding—described in one way or another—that both would have chosen. From this wider perspective, Burge's argument is unconvincing because it does not establish that it is reasonable to assume that they would have chosen different norms of meaning if they had been presented with the same alternatives.

There are two ways the externalist could attempt to respond to this objection. The first is to claim that how the person in our community and his twin would have deferred if confronted with all the relevant alternatives is not the crucial point; instead, what is decisive is the fact that they defer to their respective communities as described by Burge. The problem with this response is that it is inconsistent with analyses of meaning and understanding that assume that questions about concept possession must be answered on the basis of dispositional considerations. To be sure, dispositional accounts of meaning and understanding have been developed in different ways (Kripke, 1982; Guttenplan, 1994; McManus, 2000; Kusch, 2005), but for the purposes here these further differences are not crucial. My aim is merely to show that Burge's argument is unconvincing, and for this purpose it is sufficient that considerations about how a person is disposed to apply a term in counterfactual contexts seem relevant for questions of understanding. There is an immediate, intuitive appeal to the idea that a person's overall dispositions to apply a term are relevant for determining which concept he expresses. If we can form a clear view of how a person understands a term not only in actual cases, but also in

relevant counterfactual ones, then it seems reasonable to assume that his concept has an extension that matches both the actual and counterfactual applications.[7]

This idea has been central in very many philosophical analyses of concept possession, and not only classical accounts of straightforward one-to-one applications of natural kind concepts. To see this, we may consider again Peacocke's above analysis of implicit conceptions of a concept like the limit of a series. For Peacocke, the crucial point is that if a person who has an implicit conception of this concept were confronted with its correct explanation and, given time to consider it, then he would defer to this explanation and thereby show that our normative concept is the one that best matches his understanding. Peacocke (1998) thinks that this idea is plausible in itself, but we need not accept Peacocke's specific analysis of implicit conceptions in order to accept the more general idea that counterfactual considerations about how a person is disposed to use a concept should have an intuitive role in theories of concept possession. The problem is that Burge does not confront this idea in his arguments.

The other way the externalist can attempt to respond to the dispositional challenge is to argue that even though a dispositional approach to concept possession is correct, the person in our community and his twin would not have accepted and deferred to the same explanation of the term in question. To this response I emphasize that I have not claimed that they would have deferred to the same meaning. For the purpose of showing that Burge's argument is unconvincing, I merely needed to point to the fact that it is not reasonably clear that they would have deferred to different meaning explanations. It is simply difficult to imagine the whole scenario, and especially difficult to determine what the twin would have done. We do not seem to have any robust intuitions about this.

This is a general point that applies no matter how the thought experiment is formulated. Consider, for instance, the suggestion that the person in our community would defer to our understanding (and the twin to the counterfactual community) simply because it is his actual community. One might seek to ground this suggestion in general ideas about how a person can feel attached psychologically to things in the actual world, so that he would choose to have these things because they are the things that they actually are. An analogy can make this idea clearer: a person might be attached to his actual old car in a way such that he would not prefer to have a new car even though the old car is noisy, cramped and unreliable.

So why not suppose that the person in our community would prefer to defer to *our* understanding for the same kind of reason?

[7]There are well known challenges related to the need of specifying cases of application that determine the nature of a person's concepts (Loewer, 1997; McManus, 2000; Kusch, 2005). For instance, the concept I express by "dog" can be *dog* even if I mistakenly apply "dog" to a cat that from a long distance looks to me like a dog. But the problem of specifying content-determining applications in a non-circular way does not undermine the general idea that a person's concepts are individuated both on the basis of relevant actual and counterfactual behavior. It is this general idea that Burge does not confront.

[8]What if Burge simply assumed that the persons in the thought experiment would choose their respective actual communities? There are two obvious problems with this strategy. First, if Burge's premises are either questionable or insufficiently justified, then social externalism based on partial understanding is not a very interesting philosophical theory. Secondly, if social externalism based on partial understanding is going to describe a widespread phenomenon (that persons' concepts are individuated "externalistically") as Burge claims that the position does, then the assumptions Burge makes must correspond to how persons with a partial understanding would tend to behave in real life.

According to this objection, the person in the "mortgage" case (2 above) might not defer to an explanation of "mortgage" that represents our understanding if he does not know that it is our understanding. But the crucial point, some might suggest, is that he would choose our understanding if he knew that it was our understanding.

However, even if we grant to the externalist that the person has this kind of knowledge the problem remains the same: in order for Burge's argument to be "intuitively compelling" as he claims that it is, then it also has to be intuitively compelling that persons with a partial understanding feel attached to the normative meaning in their linguistic communities in the same way that we can feel attached to other things. But this is far from obvious. In the "mortgage" case, it is not clear that the person in our community case would experience a special attachment to our correct definition of "mortgage". We do not seem to be emotionally or cognitively attached to the standard meaning of language expressions in the way we are sometimes attached to actual persons or things.[8]

It is important to bear in mind that this objection does not require that we pick out the understanding that best matches that of the person with the partial understanding. It is sufficient that we can think of one alternative meaning explanation that the person might choose to defer to. In the "mortgage" case we may in fact use Burge's description of the term's meaning in the counterfactual community as an alternative explanation. As shown above, Burge imagines "our man constant in the ways previously indicated and that 'mortgage' commonly applied only to mortgages on houses" (Burge, 1979: p. 83).

The difference between this alternative explanation and our standard understanding is that our standard concept applies not only to loans to finance the purchase of houses, but also to loans to finance the purchase all forms of real estate (including property like land). As Burge characterizes the first step of the thought experiment, this idea about real estate is very remote from the person's understanding. So suppose that the person was asked to choose between the alternative explanation and our standard understanding. It does not seem unreasonable to assume that he would reason as follows: "This alternative way of understanding 'mortgage' seems to capture my understanding better than the comprehensive idea about real estate. I had absolutely no idea that it is possible to have a 'mortgage' on land. So it seems that my concept of mortgage, as I have understood it, corresponds to this more narrow way of understanding the term."

Of course, the person might *then* choose to change his understanding so that it matches our standard understanding, but that would not help the social externalist. The reason is that the externalist is committed to making a distinction between a person who defers to an explanation of a word on the basis of an understanding he has, and a person who defers simply because he thinks that an explanation is correct (regardless whether or not he has a partial understanding). Consider as an example of the latter case a person who is beginning to learn English and who thinks for some time that "blue" means *red*. When he discovers his mistake he will defer to the public meaning of "blue" and change his beliefs about the term, but that does not mean that he earlier expressed the concept *blue* by "blue".

This is obviously part of the reason why Burge requires that the person in the thought experiment has a minimal understanding. Burge's idea is that a person with a minimal understanding has our concept if he is deference-willing and if his

understanding matches our standard reasonably well. The problem for Burge is that even if we grant that these two conditions are met in the "mortgage" case, it is not unreasonable to assume that the person would think that the alternative explanation *better* matches his understanding and that he therefore would think that his *present* concept corresponds to this explanation.

We can strengthen this objection further by imagining that the person was confronted by a group of speakers who understood "mortgage" in the alternative way. The suggestion would then be that this would make him even more confident that he has an alternative concept, that the fact that a group of speakers has a similar understanding would make it even more probable that he would think that his use of term belongs within what Wittgenstein would call a "language-game" that does not correspond to the standard definition in a dictionary. Again, the present argument does not require that this is obvious. For the purpose of showing that Burge's argument is unconvincing, it is sufficient to point to the fact that we do not have robust intuit-tions about this. Further arguments are needed, but Burge has not provided any.

This fundamental uncertainty about the preferences of the persons in the thought experiment relates in particular to the fact that it is possible for "twin earth" thought experiments to have an important role in philosophical arguments as long as we restrict our assumptions to how the counterfactual community actually is. It is quite a different matter when we also have to determine, on the basis of these assumptions, how the twin in the counterfactual community would have acted if confronted with yet another counterfactual understanding. In order to determine this, we need to determine how the twin would have thought about himself, his understanding and his community in this even more hypothetical situation. The thought experiment becomes so complex that it seems overwhelmingly difficult to make robust interpretations without ramifying further substantial and theoretical assumptions.

This issue is not at all confronted in Burge's arguments for social externalism based on partial understanding. What Burge appeals to in his defense of the third step of the thought experiment is the idea that an understanding that is not coextendsional with our understanding cannot correspond to our concept: "The word 'arthritis' in the counterfactual community does not mean *arthritis*. It does not apply only to inflammations of joints" (Burge, 1979: p. 79). As shown above, critics of Burge have argued that the person with the incorrect understanding of "arthritis" in our community possesses the standard concept of the counterfactual community because it directly matches his understanding. This objection cannot succeed in cases of partial understanding, but I have argued that Burge faces a further problem related to deference-willingness: in order to make it reasonably clear that the twin would have chosen to defer to the standard in his community, then this standard must be defined to approximate his understanding. But then it is not unreasonable to assume that the person in our community would have deferred to this understanding as well.

More generally, it is not implausible that the person in our community and his twin would have deferred to the same understanding if they were presented with the same choices. And if they would, then a reasonable "wide" interpretation of the externalistic assumption about deference-willingness implies that they have the same concept. But then social externalism based on partial understanding is false. The social externalist could retreat to a "narrow" conception of deference-willingness, but at the cost of adhering to a problematic and counterintuitive conception.

Conclusion

Social externalism based on partial understanding escapes the standard objection to the argument for social externalism based on incorrect understanding. But social externalism based on partial understanding faces another problem related to the assumption about deference-willingness to normative meaning. Burge claims that the person with a partial understanding of a term in our community has our concept since he is willing to defer to our correct understanding, and that his counterfactual twin has another concept since he lives in a community where the term has another standard meaning.

I have argued that in order for this argument to be convincing, it has to be reasonably clear that deference-willingness should not be understood as a wider dispositional attitude, or that the person in our community and his twin would not have deferred to the same understanding if they were presented with the same choices. The main problem for Burge is that this is not reasonably clear. Burge rests the main part of his defense of social externalism on what he thinks of as the immediate appeal of the three steps of the externalistic thought experiment. In this way Burge accepts the widespread view that philosophical thought experiments can support substantial philosophical conclusions if we have strong intuitions about how they should be interpreted. The problem is that we do not have such intuitions about some of the crucial assumptions social externalism based on partial understanding requires. The externalist needs to present further substantial arguments for these assumptions, but there are good reasons for being skeptical about the possibility of finding such arguments.

REFERENCES

Bach, K. (1994). *Thought and language*. Oxford: Clarendon Press.

Burge, T. (1979). Individualism and the mental. In P. French, T. Uehling, & W. Wettstein (Eds.), *Midwest studies in philosophy, 4*.

Burge, T. (1986). Intellectual norms and foundations of mind. *The Journal of Philosophy, 86,* 3-45.

Burge, T. (1989). Wherein is language social? In A. George (Ed.), *Reflections on chomsky*. Oxford: Basil Blackwell.

Crane, T. (1991). All the difference in the world. *The Philosophical Quarterly, 41,* 1-25. doi:10.2307/2219783

Cummins, R. (1989). *Meaning and mental representation*. Cambridge, MA: MIT Press.

Evans, G. (1982). *The varieties of reference*. Oxford: Clarendon Press.

Fodor, J. (1998). *Concepts*. Oxford: Oxford University Press. doi:10.1093/0198236360.001.0001

Guttenplan, S. (Ed.) (1994). *A companion to the philosophy of mind*. Oxford: Basil Blackwell.

Higginbotham, J. (1998). Conceptual competence. In E. Villanueva (Ed.), *Concepts: Philosophical Studies, 9*.

Jacob, P. (1987). Thoughts and belief ascriptions. *Mind & Language, 4,* 301-325. doi:10.1111/j.1468-0017.1987.tb00124.x

Kripke, S. (1982). *Wittgenstein on rules and private language.* Oxford: Blackwell.

Kusch, M. (2005). Fodor v. Kripke: Semantic dispositionalism, idealization and ceteris paribus clauses. *Analysis, 65,* 156-163. doi:10.1093/analys/65.2.156

Loar, B. (1985). Social content and psychological content. In R. Grimm, & D. Merill (Eds.), *Contents of thought*. Tucson, AZ: Arizona Uni-

versity Press.

Loewer, B. (1997). A guide to naturalizing semantics. In B. Hale, & C. Wright (Eds.), *A companion to the philosophy of language*. Oxford: Basil Blackwell.

McManus, D. (2000). Boghossian, Miller and Lewis on dispositional theories of meaning. *Mind & Language, 15,* 393-399. doi:10.1111/1468-0017.00141

Nordby, H. (2004). Concept possession and incorrect understanding. *Philosophical Explorations, 7,* 55-70. doi:10.1080/1386979032000186854

Peacocke, C. (1992). *A study of concepts*. Cambridge, MA: MIT Press.

Peacocke, C. (1998). Implicit conceptions, understanding and rationality. In E. Villanueva (Ed.), *Concepts: Philosophical, 9*.

Pessin, A., & Goldberg, S. (Eds.) (1996). *The twin earth chronicles*. New York/London: M. E. Sharpe.

Putnam, H. (1975). *Mind, language and reality* (Vol. 2). Cambridge:

Cambridge University Press. doi:10.1017/CBO9780511625251

Putnam, H. (1996). Introduction. In A. Pessin, & S. Goldberg, (Eds.), *The twin earth chronicles*. New York/London: M. E. Sharpe.

Rey, G. (1998). What implicit conceptions are unlikely to do. In E. Villanueva (Ed.), *Concepts: Philosophical Studies, 9*.

Schantz, R. (Ed.) (2004). *The externalist challenge*. Berlin/ New York: Walter de Gruyter.

Schiffer, S. (1998). Doubts about implicit conceptions. In E. Villanueva (Ed.), *Concepts: Philosophical Studies, 9*.

Toribo, J. (1998). The implicit conception of implicit conceptions. In E. Villanueva (Ed.), *Concepts: Philosophical Studies, 9*.

Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Basil Blackwell.

Woodfield, A. (1982). Thought and the social community. *Inquiry, 25,* 435-450. doi:10.1080/00201748208601979