CONCEPTUAL ERRORS AND SOCIAL EXTERNALISM

By Sarah Sawyer

Åsa Maria Wikforss has recently proposed a novel individualist response to Burge's thought-experiments in favour of social externalism. Her response allows the individualist to maintain that narrow content is truth-conditional without being idiosyncratic. The narrow aim of this paper is to show that Wikforss's argument against social externalism fails, and hence that the individualist position she endorses is inadequate. The more general aim is to attain clarity on the social externalist thesis. Social externalism need not rest, as is typically thought, on the possibility of incomplete linguistic understanding or conceptual error. The unifying principle that underlies the various externalist thought-experiments is identified.

Social externalism is the thesis that many of a subject's mental states and events are dependent for their individuation on the subject's social environment. The thesis opens up the possibility that a subject's mental state and event kinds might vary with variations in the social environment, even while the subject's physical properties, including her functional properties and her physical history, all individualistically and non-intentionally described, remain constant. The thesis has been argued for persuasively by Tyler Burge on the basis of a series of by now well known thought experiments.[1]

In response to the thought experiments, hard-headed individualists have thought themselves forced to adopt one of two broadly defined positions. The first accepts the conclusions of the thought experiments – accepts that ordinary truth-conditional psychological content is anti-individualistically individuated – but maintains nonetheless that there is an important type of psychological content, albeit non-truth-conditional

---

[1] See T. Burge, 'Individualism and the Mental', *Midwest Studies* IV (1979) pp. 73-121; 'Other Bodies', in A. Woodfield ed. *Thought and Object* (Oxford: Clarendon Press, 1982) pp. 97-120; 'Intellectual Norms and Foundations of Mind', *Journal of Philosophy* 83:12 (1986) pp. 697-720.

content, which is individualistically individuated, and which is of interest to scientific psychology.[2] The second rejects the conclusions of the thought experiments, but at the cost of legislating systematic reinterpretations of ordinary psychological attitude ascriptions.[3] The reinterpretations may take one of a number of various forms, but each is problematic in its own right, and each involves a significant loss of generality with regard to psychological generalisations.[4] Individualists, then, seem faced with an unhappy dilemma. As Åsa Maria Wikforss has recently put it, 'they can give up on reference, thereby compromising their individualism; or they can reject the communitarian conclusions and accept conceptual and referential fragmentation'[5].

Wikforss maintains, and I agree, that neither option is palatable. In her paper, Wikforss goes on to propose a different response to the anti-individualist conclusions of the thought experiments, one which has the benefit of avoiding the unhappy dilemma altogether. This is important since if her response to the anti-individualist were sound, Wikforss would in addition have provided a novel position for the individualist to occupy, one which is clearly preferable to the individualist positions identified previously. The specific aim in this paper is, then, to examine Wikforss's 'third way'. In so doing, I aim to show that Wikforss's argument against social externalism fails, and hence that the individualist position she endorses is inadequate. The more general aim of this paper is to attain clarity on the social externalist thesis. As will emerge, social externalism need not rest, as is typically thought, on the possibility of incomplete

---

[2] For example see J. Fodor, 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology', *Behavioral and Brain Sciences* 3 (1980) pp. 63-73; C. McGinn, 'The Structure of Content', in A. Woodfield (ed.) *Thought and Object* (Oxford: Clarendon Press, 1982) pp. 207-58; and S. Stich, *From Folk Psychology to Cognitive Science: The Case Against Belief* (Cambridge, Mass.: MIT Press, 1983).

[3] For example see B. Loar, 'Social Content and Psychological Content', in H. Grimm & D. Merrill (eds) *Contents of Thought* (Arizona: University of Arizona Press, 1985); and G. Segal, *A Slim Book about Narrow Content* (Cambridge, Mass.: MIT Press, 2000).

[4] Burge has argued at length against a wide variety of reinterpretation strategies, together with the apparent motivations for each, in 'Individuation and the Mental', esp. § III.

[5] A. M. Wikforss, 'Social Externalism and Conceptual Errors', *The Philosophical Quarterly* 51: 203 (2001) pp. 217-31, at p. 218.

linguistic understanding or conceptual error.[6] A variety of quite distinct thought experiments have been offered in favour of social externalism, and it is important to see both that there is a unifying principle which underlies them, and also what that principle is.

I. THE FIRST KIND OF THOUGHT EXPERIMENT: INCOMPLETE UNDERSTANDING

In 'Individualism and the Mental' we are offered the following three-step thought experiment in support of social externalism. We are to suppose first that a subject S, who is rational, intelligent, and generally competent in English, has a large number of beliefs commonly attributed with content clauses containing the word 'arthritis' in oblique occurrence. Many such of S's beliefs are true, but S also believes falsely that he has developed arthritis in the thigh. S discovers that his belief is false when his doctor tells him that arthritis is specifically an inflammation of the joints and that any dictionary could have told him the same. S relinquishes his belief. The second step consists of a counterfactual supposition in which S's physical history and intentional phenomena, non-individualistically described, are assumed to be the same as in the original situation up until the time at which S expresses his fear to his doctor. The counterfactual difference concerns the word 'arthritis', which in its counterfactual use is commonly applied, and defined to apply, not only to arthritis but also to various other rheumatoid ailments outside the joints. The word 'arthritis' in the counterfactual situation differs in dictionary definition and in extension from the word 'arthritis' in the actual situation. The final step is an interpretation of the counterfactual situation. Given that the word 'arthritis' in the counterfactual situation does not mean *arthritis*, it is plausible to suppose that in the counterfactual situation S lacks thoughts involving the notion *arthritis*. The truth-values of the subjects' thoughts differ. If the interpretation is correct, it is possible for mental content to vary solely with variations in the social environment. (As Burge is keen to emphasise, the thought experiments are not deductive arguments. In particular, there are

---

[6] This claim has already been made by Burge in 'Intellectual Norms and Foundations of Mind'. However, given the almost exclusive focus in the literature on his 'Individuation and the Mental', it is worth restating.

alternative interpretations of the counterfactual situation which are consistent with individualism. However, the anti-individualist conclusions are highly plausible, and the individualist interpretations come at a price, viz. the price of widespread reinterpretation of subjects' utterances).

Wikforss maintains that Burge's anti-individualism relies essentially on the idea that it is possible for a subject to make a conceptual error, without her ability to think with the incompletely understood concept thereby being undermined. This is not an uncommon assumption. Several passages from Burge provide prima facie support for this claim. Thus,

> The argument can get under way in any case where it is intuitively possible to attribute a mental state or event whose content involves a notion that the subject incompletely understands. As will become clear, this possibility is the key to the thought experiment. ('Individualism and the Mental', p. 79).

> *[I]f* the thought experiment is to work, one must at some stage find the subject believing (or having some attitude characterized by) a content, despite an incomplete understanding or misapplication. An ordinary empirical error appears not to be sufficient (p. 83).

However, if we look more closely we see that these and other passages in which Burge makes explicit reference to the key notions of incomplete understanding and conceptual error are concerned not with anti-individualism per se, nor even with social externalism per se, but rather with the particular kind of thought experiment under discussion at the time. The passages would establish that Burge sees social externalism as resting on the notions of incomplete understanding and conceptual error only if this were the only kind of thought experiment available to support social externalism. This is not the case, as we can see by examining a thought experiment of a different kind put forward by Burge in a later paper ('Intellectual Norms and Foundations of Mind').[7]

---

[7] I leave open the possibility that Burge, at the time of writing 'Individualism and the Mental', took anti-individualism to rest essentially on the notion of incomplete understanding.

II. THE SECOND KIND OF THOUGHT EXPERIMENT:
NONSTANDARD THEORY

We are to suppose first that a subject *A*, who is rational, intelligent, and generally competent in English, has a large number of beliefs commonly attributed with content clauses containing the word 'sofa' in oblique occurrence. *A* can use the term 'sofa' reliably and picks up the normal truisms. However, *A* at some point comes to doubt the truisms concerning sofas and hypothesises that sofas are works of art or religious artifacts, not pieces of furniture to be sat upon. Nevertheless, *A* is willing to test his hypothesis empirically and the tests he proposes are reasonable. The second step of the thought experiment consists of a counterfactual supposition in which *A*'s physical history and intentional phenomena, non-intentionally described, are assumed to be the same as in the original situation. In the counterfactual situation the objects that look like sofas are, and are widely known to be, works of art or religious artifacts that would typically collapse under a normal person's weight. There are no sofas in the counterfactual situation, and the word form 'sofa' differs in dictionary definition and in extension from the word 'sofa' in the actual situation. Call the subject in the counterfactual situation '*B*', and the objects 'safos'. *B* assumes that most people believe these objects to be pieces of furniture. But, like *A*, he begins to develop doubts. By the time *B* expresses his scepticism he correctly doubts that safos function as pieces of furniture to be sat upon. The conclusion is that *A* and *B* are physically indistinguishable but they have different mental states. *A* has false beliefs about what sofas are, but correct beliefs about the sociolinguistic practice that surrounds him; *B* has true beliefs about safos but false beliefs about his surrounding sociolinguistic practice.

Drawing a distinction between three kinds of thought experiment, and emphasising the second discussed here, Burge writes, ('Intellectual Norms', pp. 708-9).

> The arguments of "Individualism and the Mental" and "Other Bodies" … ascribe incomplete linguistic understanding and ignorance of expert knowledge (respectively) to the relevant protagonists. By contrast, *A* may be a sophisticate. He need not lack linguistic understanding or be unapprised of expert or common opinion. The present argument features not incomplete understanding or ignorance of specialized knowledge, but nonstandard theory.

Burge goes on to emphasise (p. 709) that this second kind of thought experiment can be 'adapted to any substantive notion that applies to physical objects, events, stuffs, properties', thus having an extremely wide application, since any such notion can be subject to nonstandard theorising. The distinction between the thought experiments is crucial to bear in mind given the almost exclusive focus in the literature on the argument presented in 'Individualism and the Mental', and given also that the Achilles' heel in Wikforss's argument, as we will see later, is the reliance on thought experiments driven by incomplete understanding.

The following two points are essential to the second kind of thought experiment. First, the error involved is not one of linguistic misunderstanding, nor is it best thought of as conceptual: *A* is not confused or mistaken about whether the objects he sees are sofas. Rather, the error is largely empirical: *A*'s doubts concern *whether sofas are what people think they are* (p. 711). Similar considerations apply to *B*. Second, the argument rests on the possibility of being able coherently to doubt what Burge calls 'meaning-giving normative characterisations'. Normative characterisations are statements that purport to give basic and necessarily true information about what certain things are. Some such normative characterisations are meaning-giving. Thus 'sofas are upholstered pieces of furniture for two or more people … used for sitting on' approximates a meaning-giving normative characterisation for sofas. Such meaning-giving normative characterisations are grounded in use in the sense that they are reached by a dialectic process which proceeds by reflection on archetypical examples. When the most competent speakers reach agreement on a characterisation, the characterisation correctly specifies the meaning of the relevant linguistic expression. Thus the meanings of linguistic expressions are fixed by use and by agreement. Of course, the dialectic process need not, and typically will not, take place explicitly. Rather, the linguistic meaning of a given expression must accord with the use of that expression by those widely regarded as the most competent speakers in the community, and must treat the characterisations that such speakers would give as 'at least approximations to the norm' ('Intellectual Norms', pp. 703-4).

However, despite the fact that linguistic meaning is given by reflective agreement, the role of examples in the dialectic ensures that the concept shared by the speakers can correct the meaning-giving characterisations those speakers alight on. This means that it is possible for thought to correct linguistic meaning. This is because the explication of the linguistic meaning of an expression need not provide correct application conditions for the associated concept. It also means that it is possible for a fully competent speaker coherently to doubt a meaning-giving normative characterisation. The relevant doubt would amount to no more than wondering whether the items concerned had indeed been correctly characterised, and it need not reflect anything that would standardly be considered a conceptual error on the part of the subject. This remains true even though, as Burge emphasises, there is no clear divide between empirical truths and conceptual truths, or truths of meaning. The fact that there is no such divide can be clearly seen by the fact that determining what Xs are and determining what 'X' means are part of the same project. The characterisations are, after all, *empirical* claims, and yet they are also *meaning*-giving.

Since synonymy statements are also grounded in use - a statement of the form 'Fs are Gs' is a synonymy statement if and only if the most competent users of the terms would use 'F' and 'G' interchangeably – it is also possible coherently to doubt synonymy statements. Indeed, I take it that this is precisely what A is doing, since he doubts that sofas are upholstered pieces of furniture for two or more people … used for sitting on. [8]

III. WIKFORSS'S ARGUMENT: TWO PROBLEMS

Wikforss discusses the analysis of meaning-giving normative characterisations given in 'Intellectual Norms', but her argument nevertheless remains focused on the notions of conceptual error and incomplete understanding that appear in 'Individualism and the Mental'. Correspondingly, her focus is on the first kind of thought experiment rather than

---

[8] Burge was clear about the possibility of coherently doubting a synonymy statement when he wrote 'Individualism and the Mental'. He had earlier argued for the possibility in his 'Belief and Synonymy', *Journal of Philosophy* 75:3 (1978) pp. 119-38, and he explicitly endorses the possibility again in his 'Intellectual Norms'.

the second, despite the fact that it is the second, and not the first, that makes essential play with the notion of a meaning-giving normative characterisation. Focusing entirely on the first kind of thought is, I believe, the root of the trouble with Wikforss's argument against social externalism. We are now in a position to discuss her argument directly.

Wikforss focuses on the claim that meaning-giving normative characterisations may be false. This claim follows from the fact that such characterisations are arrived at by reflective agreement. Clearly what the most competent speakers agree Xs are like may nevertheless fail to capture what Xs are in fact like. The characterisations may themselves reflect incorrect theory. Wikforss concludes from the gap between the nature of the items being characterised and their purported nature as specified by the characterisations that a subject may consistently doubt the experts' characterisations while nevertheless having all true beliefs about Xs. That is, in certain cases the doubting subject may in fact be in possession of the correct theory about Xs. On this much we can agree.

Wikforss goes on to apply this result to the original thought experiment as follows. In the actual situation, 'arthritis afflicts the joints only' is a meaning-giving normative characterisation agreed on by the experts and central to the meaning of the term 'arthritis'. However, the fact that S does not accept the claim embodied in the characterisation is consistent with his having a complete understanding of the concept *arthritis*, and hence consistent with his making no conceptual error. After all, the characterisation, and hence the claim S rejects, could be false. But then there is no reason to think that the term 'arthritis' in the counterfactual situation expresses a different concept from the term 'arthritis' in the actual situation. The two communities could simply have different theories about the same disease. Wikforss writes (p. 225) 'Indeed, the counterfactual community could as well be our future community, one in which medical discoveries have made the experts conclude that arthritis can afflict the ligaments as well as the joints'.

Wikforss's argument is important, since if her response to the anti-individualist were sound she would have provided a novel position for the individualist to occupy. The individualist, she maintains, can grant that S has the standard concept *arthritis*, but since any mistake S makes need not be thought of as conceptual, the individualist can deny that S's concept in the actual situation differs from the concept he possesses in the

counterfactual situation. Wikforss's 'third way', then, depends upon taking seriously Quine's idea that what might be taken as a definitional or central truth at one time can be rejected without a change in meaning. Wikforss appeals to Burge's account of meaning-giving normative characterisations to support her point: if the subject in either the actual or the counterfactual situation is guilty of making a mistake, it need not be considered a conceptual error or an error in linguistic understanding.

There are two problems with Wikforss's argument. First, it is ineffective in its own right. This is because, on reflection, Wikforss's argument is an epistemic argument rather than a metaphysical one. We can grant first, that for all Alf and his cohorts know arthritis can occur outside the joints. That is, we can grant that Alf's characterisation of arthritis may be correct. Second, we can grant that in certain cases what appears to be a conceptual error on the part of one subject is in fact an empirical error on the part of another. But neither of these points serves to undermine the first kind of thought experiment in support of social externalism. Burge has offered us a thought experiment in which it is stipulated that Alf has a false belief about arthritis, since it is stipulated that 'arthritis' is correctly applied to ailments within joints only. The thought experiment is Burge's, and we surely cannot deny him this stipulation. It would seem that Wikforss has confused the position of the subject with the position of the theorist. When we imagine ourselves as Alf's cohorts, we imagine ourselves in a position from which we cannot tell whether Alf has a false belief or whether we have a false theory. Burge's stipulation rules out the latter as an option. In addition, cases such as the one Burge describes surely sometimes arise.

This point can be clarified as follows. For any given word there will typically be numerous subjects each of whom would apply that word in slightly different ways in non-central cases. But we would still in many of these cases attribute the standard concept expressed by that word to each of them. An ability to discriminate some but not all Fs from non-Fs typically belies a partial grasp of the concept *F*. If a subject were to apply the term 'sofa' to overstuffed armchairs as well as to sofas, for example, it would be most plausible to attribute her an incomplete grasp of the concept *sofa*. This kind of error is surely best thought of as conceptual. In contrast, an ability to discriminate Fs from non-Fs combined with an inability correctly to characterise the nature of Fs typically belies

incorrect empirical theory. Thus in the second thought experiment above, *A* is able reliably to distinguish sofas from non-sofas, but characterises the nature of sofas incorrectly. In this case it is most plausible to attribute *A* a full grasp of the concept *sofa*, but an incorrect empirical theory of sofas. That some cases fall into the latter category suffices to demonstrate that Wikforss's objection fails. One important distinction to emerge from this discussion is that between a concept and its explication. A subject's grasp of a concept is tied to her ability to apply the concept correctly. It is possible for a subject to grasp a concept fully, however, without being able to give a correct explication of the concept, that is, without being able to articulate a correct meaning-giving normative characterisation for the associated word.

I shall turn now to the second, related problem with Wikforss's argument: it will not work against the second kind of thought experiment. (I do not mean to imply that Wikforss intended her argument to have such a wide application.) This is because the second kind of thought experiment in no sense relies on the notion of incomplete understanding. Let me explain. In the second thought experiment, *A* and *B* are assumed to be in environments that differ in the following respects. First, the linguistic meanings of their respective terms 'sofa' differ. Second, and crucially, the actual situation is assumed to contain sofas, whereas the counterfactual situation is assumed to contain *no* sofas. The objects that look like sofas in the counterfactual situation are explicitly assumed to be works of art or religious artifacts, and it is assumed that most of them would collapse under the weight of a normal person. Since the referent plays a role in the individuation of the concept, it is this second assumption that generates the externalist conclusion that the concepts of the subjects differ. It also entails that *A*'s community and *B*'s community do not have different theories about the same things, but have, rather, different theories about *different* things. It is important to note that the assumption does not rely on the truth of any of our actual meaning-giving normative characterisations. The thought experiment could just as well be run by appeal to two non-actual situations. In the thought experiment as Burge sets it up, *A* proposes reasonable empirical tests to determine whether his doubts about the meaning-giving normative characterisations are correct, and it is legitimate to suppose that once the tests are carried out, *A* and the rest of his community are satisfied, and correct in collectively rejecting his doubts.

IV. THE UNIFYING PRINCIPLE OF EXTERNALISM

The crux of externalism lies with the claim that referents themselves play a role in the individuation of concepts. Concepts are individuated partly by their referents rather than entirely by what the subject thinks is true of the referents. (The so-called 'empty case', in which a concept lacks a referent, poses special problems with which I will not deal here.) The various thought experiments are designed to bring out our commitment to this claim as it is evidenced by our ordinary practice of psychological state attribution. *This claim is the principle that unifies the thought experiments.* One way to bring out our commitment to the claim is to focus on cases of incomplete understanding and conceptual error. But it is not the only way. Crucially, if the nature of the referents can trump what is thought to be true of the referents, then the following possibility opens up: what is thought to be true of a set of objects or properties in one situation can in fact be true of a different set of objects or properties in a different situation. Again, it is the nature of the referents and the role they play in concept-individuation that is of primary importance. The notion of conceptual error is secondary and derivative.
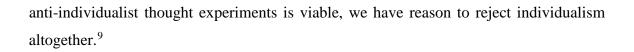
Wikforss states (p. 225) that her arguments do not work against physical externalism on the grounds that 'physical externalism can get by without appealing to conceptual truths'. In this paper my primary concern has been to urge that social externalism can also get by without appealing to conceptual truths. This is, so I have argued, shown by the kind of thought experiment that appeals to nonstandard theorising. But this raises a residual worry. On reflection, one might begin to wonder whether this second kind of thought experiment establishes a form of *social* externalism at all. After all, the differences in mental content as I have described the situation are ultimately due to differences in the *physical* environment. Indeed, Burge has claimed that these considerations show that social factors *cannot* be the final arbiter in the individuation of mental events. He writes,

the dubitability of meaning-giving normative characterizations can be converted into a demonstration that social practices are not the only or ultimate non-individualistic factor in

individuating mental states and events. … even where social practices are deeply involved in individuating mental states, they are often not the final arbiter. This is because the sort of agreement that fixes a communal meaning and norms for understanding is itself, in principle, open to challenge ('Intellectual Norms', p.707).

However, the apparent worry is removed by focusing on the following consideration. The reason Wikforss is concerned specifically with social externalism is in large part because of its extremely wide application. But Burge's claims in the passage just quoted do nothing to change the breadth of application of the externalism under consideration, whether it be construed as social externalism or, more properly speaking, broad physical externalism – 'broad', since the physical differences to which Burge alludes are clearly differences not in natural kinds, but rather in the physical constitution and history of the objects involved, and in the uses to which each community puts such objects. Wikforss's position is in trouble either way.

Wikforss's proposal is to reject the assumption that she takes Burge's thought experiments to depend upon: namely, the assumption that the subjects are guilty of incomplete linguistic understanding or conceptual error. Rejecting this assumption, she hopes, will provide an alternative to the unhappy horns of the standard individualist dilemma, and hence provide a motivation for her own more palatable form of individualism. However, as we have seen, not all of Burge's thought experiments depend upon this assumption. Rather, the thought experiments are designed to bring out, in one way or another, our commitment to the claim that referents serve partly to individuate concepts. Consequently, rejecting the assumption will not suffice as a response to the thought-experiments. A widespread form of externalism can be established even if it is granted that conceptual error and linguistic misunderstanding are rare indeed. (I do not think it plausible to grant either of these claims. However, I am here concerned to argue simply that social externalism does not depend upon such error or misunderstanding.) If Wikforss is right to maintain that neither of the standard individualist responses to the

anti-individualist thought experiments is viable, we have reason to reject individualism altogether.[9]

University of Kansas