

Deference and Self-Knowledge

It has become increasingly popular to suggest that non-individualistic theories of content¹ undermine our purported *a priori* knowledge of such contents because they entail that we lack the ability to distinguish our thoughts from alternative thoughts with different contents. However, problems relating to such knowledge of ‘comparative’ content tell just as much against individualism as non-individualism. Indeed, the problems presented by individualistic theories of content for self-knowledge are at least, if not more, serious than those presented by non-individualistic theories. Consequently, considerations of self-knowledge give one no reason to embrace individualism. If anything, they give one reason to reject it.

That there should be a tension between non-individualistic theories of content and self-knowledge is not surprising. If features of our (physical and) social environment contribute to the contents of our thoughts,² and we don’t have *a priori* access to these features, then it would seem that we should lack *a priori* access to the content of our own thoughts.³

An influential reply to this sort of worry was provided by Tyler Burge, who stressed that our judgments about our thoughts have their contents determined by the very same external factors that determine the content of our ‘first-level thoughts’. Consequently, ‘second level’ thoughts such as “I am thinking that water is wet” will always be true.⁴ Since the very same factors that determine the contents of our thoughts also determine the content of our judgments about these thoughts, the fact that we may not be authoritative about these factors does not prevent our judgments about the contents of our thoughts from being authoritative.

¹ Such theories are often also referred to as “externalistic” or “anti-individualistic”. My reasons for preferring “non-individualistic” can be found in Jackman 1996, 1998.

² As suggested in Putnam 1975 and Burge 1979.

³ See Woodfield 1982, p. viii.

⁴ “Knowledge of one’s own mental events ... consists in a reflexive judgment which involves thinking a first-order thought that the judgment itself is about. The reflexive judgment simply inherits the content of the first order thought. Consider the thought, ‘I hereby judge that water is a liquid.’ What one needs in order to think this thought knowledgeably is to be able to think the first-order, empirical thought (that water is a liquid) and to ascribe it to oneself, simultaneously. Knowing one’s thoughts no more requires separate investigation of the conditions that make the judgment possible than knowing what one perceives.” (Burge 1988, 118 (See also Davidson 1987.))

⁵ See Flavey and Owens 1994, Butler 1997, Boghossian 1989.

However, even if we did have this type of ‘reflexive’ knowledge of content, there is still a type of self-knowledge that the non-individualist can’t claim that we have. Namely, the non-individualist cannot claim that we have knowledge of *comparative* content.⁵ Even if we know that our thoughts about our current thoughts must be true, we still may not know that we are having, say, water-thoughts *rather than* twater-thoughts.⁶ Without such knowledge of ‘comparative content’, it might seem as if we can’t know what we are thinking. As Brueckner puts it:

I claim to know that I am thinking that some water is dripping. If I know that I am thinking that some water is dripping, then I know that I am not thinking, instead, that some twater is dripping. But I do not know that I am not thinking that some twater is dripping, since, according to externalism, if I were on twin earth thinking that some twater is dripping, things would seem exactly as they now seem (and have seemed). So I do not know that I am thinking that some water is dripping.⁷

However, to know what our thoughts are, we need not distinguish them from *every possible* thought that we might have in alternate environments. The fact that I could not distinguish my parents from their twin earth counterparts (were such counterparts to exist and be switched with my parents), does not change the fact that I know my parents when I see them. Such possible counterparts are not alternatives that I need to rule out when I claim to know that, say, my parents are in the room with me. In much the same way, possible twater-thoughts are not the sorts of things I need to rule out when I claim to know that, say, I am thinking that some water is dripping. One need only rule out relevant alternatives, and since there is no twin earth (or if there is one, since we have no contact with it), twater thoughts are not relevant alternatives of the sort that would undermine self-knowledge. Knowledge of comparative content does not, then, seem necessary for knowledge of content.⁸

⁶ That is, the thoughts one would have had one acquired one’s language on ‘Twin Earth’ (a planet just like ours other than that the ‘water’ is made up of a chemical compound XYZ rather than H₂O (See Putnam 1975)).

⁷ Brueckner 1990, p. 448. See also:

The idea is that a pair of internally identical twins on earth and twin earth respectively have different but, as far as *they* can tell, indistinguishable substances, which they both call “water,” in their environments; so, given externalism, they have different ‘water’-concepts and ‘water’-thoughts. Since they have different thoughts without really being able to tell the difference, they do not fully know what their own thoughts are. (Bilgrami 1992b, 235.)

⁸ See Flavey and Owens 1994. Boghossian 1989 also endorses such a response to the suggestion that the mere possibility of, say, someone having twater thoughts, is enough to undermine self knowledge.

Paul Boghossian, however, has provided an argument that suggests that a lack of knowledge of comparative content can undermine our knowledge of content.⁹ For instance, non-individualism would seem to lead to the following sort of problem case. Assume that Smith is an inhabitant of Earth, and thus that his ‘water’ thoughts are about *water*. If Smith is then moved to Twin Earth without his knowledge, then after a number of years his ‘water’ thoughts will come to be about *twater*.¹⁰ Even if a number of such switches are made, his judgments about his *current* thought contents will always be self-verifying in the way that Burge suggests. Nevertheless, if he is informed of the switches, but not of the precise times of their occurrences, he will not be able to tell whether last year’s ‘water’-thoughts were *water*-thoughts or *twater*-thoughts. The significance of this, Boghossian claims, stems from the following “platitude about memory and knowledge”:

If S knows that p at t1, and if at (some later time) t2, S remembers everything S knew at T1, then S knows that p at t2. Now let us ask: *why* does S not know today whether yesterday’s thought was a *water* thought or a *twater* thought? The platitude insists that there are only two possible explanations: either S has forgotten or he never knew.¹¹

Since it is extremely implausible to think that Smith has forgotten anything in such a case, Boghossian concludes that, in spite of the self-verifying character of his self-ascriptions, Smith does not know the content of his ‘water’ thoughts.

There has been considerable controversy over just what Boghossian’s argument shows. Many admit that in cases such as Smith’s, where ‘switching’ *actually* occurs, *twater*-thoughts are a relevant alternative to *water*-thoughts, and so an inability to rule them out would undermine self-knowledge. Non-individualism would thus seem to rule out our *necessarily* knowing the contents of our own thoughts. This is, itself, a non-trivial conclusion that would rule out standard ‘Cartesian’ conceptions of our knowledge of our minds. Nevertheless, many insist that such ‘slow-switching’ cases rarely, if ever, occur, and when they don’t, the alternate contents still are

⁹ Boghossian 1989.

¹⁰ This assumption will be accepted here for the sake of argument, but it strikes me as implausible to suppose that natural kind terms ‘switch’ in this fashion (rather than becoming more general terms that can be applied in both environments). In this sense they are not like proper names, which are less well suited for this sort of generalization of their extension.

¹¹ Boghossian 1989, pp.171-2. The idea that the switching could be fast enough to make problematic our connection to *yesterdays* ‘water’ thoughts may strike many as implausible, but this doesn’t affect the general

not relevant alternatives that need to be ruled out.¹² Non-individualism might thus still be compatible with our *actually* having *a priori* knowledge of the contents of our thoughts. Others, however, have argued that (especially if one accepts the sorts of non-individualistic content attributions associated with Burge's work) 'switching' cases are fairly common. Peter Ludlow, for instance, argues that, if the non-individualist is right, the contents of our thoughts may change without our noticing whenever we move between linguistic (sub)communities who attach different meanings to the same terms. Such movement may be fairly common, and so switching cases could present relevant alternatives that *actually* undermine self-knowledge.¹³

Indeed, it has been suggested that, switching cases are prevalent enough for the 'switched thoughts' to be relevant alternatives even in those cases where no switching occurs.¹⁴ If counterfeit coins become generally prevalent they are a relevant alternative to real coins even in contexts where there are no counterfeit coins actually present.¹⁵ In much the same way, if our concepts 'switch' frequently, then 'switched' alternatives are relevant even for those concepts that have never switched. To know the content of *any* of our thoughts we would have to rule out the possibility of their involving switched concepts, and that is not something that we can do *a priori*. Non-individualistic accounts of content would thus lead not only to *possible* failures of self-knowledge, but also to our *actually* lacking *a priori* knowledge of our own thought contents. The purported prevalence of slow switching suggests that without knowledge of *comparative* content, knowledge of content is undermined.

Nevertheless, even if the prevalence of such unperceived content switches is enough to undermine self-knowledge, this does not give one reason to be an individualist rather than a non-individualist about content. After all, the possibility that the contents of a speaker's thoughts could

point. Of course some may reject the dilemma by questioning Boghossian's assumptions about how memory should be treated if one is a non-individualist. (See, for instance, Ludlow 1996, Brueckner 1997.)

¹² See, for instance Warfield 1992.

¹³ Ludlow 1995. See also Butler 1997.

¹⁴ As Butler puts it, "circumstances that tend to occur give rise to relevant alternatives, even if they do not actually occur in a particular instance, simply because their general prevalence indicates that they might easily have occurred." (Butler 1997, p. 780.)

¹⁵ See Ludlow 1995, p. 226-7.

change without the speaker being aware of it is hardly a problem that is unique to non-individualistic accounts of content. Indeed, quite the opposite is the case. For instance, consider the following quotation from Boghossian about a speaker, S, who believes that he has always lived on earth, but who has undergone a series of slow switches between earth and twin earth.

If someone were to ask him, just after one set of twin-earthian concepts has been displaced by a set of earthian ones, whether he has recently thought thoughts involving an arthritis-like concept distinct from *arthritis*, S would presumably say “no.” And yet, of course, according to the anti-individualist story, he has. His knowledge of his own past thoughts seems very poor, but not presumably because he simply can’t *remember* them. Could it be because he never knew?¹⁶

Boghossian’s point may be telling, but one could ask precisely the same question about what the individualist must say when Bert defers to his doctor’s use of “arthritis.”¹⁷ The non-individualist allows that Bert’s arthritis-concept need not change when his doctor corrects his belief that people can get arthritis in their thighs. The individualist, on the other hand, typically insists that what Bert’s concept has changed when he accepts the correction. Bert now, like us, has an *arthritis*-concept. However, before being corrected, he had a *tharthritis*-concept (*tharthritis* being much like arthritis, but also occurring in the thigh).¹⁸ This leaves the individualist open to precisely the sorts of rhetorical questions raised by Boghossian above. After all:

If someone were to ask Bert, just after his idiosyncratic ‘arthritis’ concept has been displaced by the standard one, whether he has recently thought thoughts involving an arthritis-like concept distinct from *arthritis*, he would presumably say “no.” And yet, of course, according to the individualist story, he has. His knowledge of his own past thoughts seems very poor, but not presumably because he simply can’t *remember* them. Could it be because he never knew?¹⁹

Such unacknowledged changes in content will be extremely common if the individualist’s account of content is correct. Indeed, one of the purported advantages of non-individualistic accounts of content is precisely that they seem to make content comparatively *stable*.²⁰

Since we typically don’t perceive ourselves as changing what we mean when we defer to correction, individualistic accounts of content suggest that our deference behavior manifests a lack

¹⁶ Boghossian, p. 160.

¹⁷ Burge 1979.

¹⁸ See, for instance, Bilgrami 1992a, p.80.

¹⁹ One might wonder, as in Boghossian’s case, whether questions about whether they were “thinking thoughts with slightly different concepts” would get any clear answer from Bert if he were not a philosopher. Nevertheless, the fact that he thinks of his last hours’ belief that he had arthritis in his thigh as *false* suggests that he is committed to the concepts being unchanged.

of self-knowledge on our part. For instance, after being corrected Bert will admit that he previously falsely believed that he had *arthritis* in his thigh. Consequently, the individualist who insists that he actually had a true belief about having *tharthritis* in his thigh will have to say that Bert is mistaken about what he believed in the past. Given that, by the individualist's lights, Bert clearly doesn't *now* know what concept he had prior to his correction, and given that it is implausible to suggest that he has forgotten anything in the interim, the individualist should conclude that Bert didn't know that he was thinking *tharthritis*-thoughts while he was thinking them. It seems, then, that whenever we (are disposed to) defer to the usage around us, the individualist must claim that we are disposed to misidentify the contents of our own thoughts.²¹ The argument for the incompatibility of non-individualism and self-knowledge presupposes that knowledge of comparative content is necessary for knowledge of content. However, since we often think that the concepts associated with our words remain unchanged in circumstances where the individualist must insist that they have changed, the individualist must deny that we have knowledge of comparative content in such cases.²²

Furthermore, the sorts of deference cases that cause problems for the individualist will be more prevalent than the switching cases that cause problems for the non-individualist. Deference seems relevant whenever we use a word in a fashion that is different from someone who's correction we are disposed to defer to. Given that we may be willing to alter our usage in deference to experts with most of our terms, it might seem as if the individualist is committed to our not knowing *any*

²⁰ For a discussion of this, see Jackman 1996, 1998.

²¹ Furthermore, things are even worse if one is committed to the sort of internalism characteristic of, say, conceptual role semantics. In that case it may seem that, in spite of their perceived constancy, the concepts associated with a term will change every time there is a change in the set of beliefs involving it. (For a discussion of this, see Fodor & Lepore 1992.)

²² It should not be surprising, then, that defenders of individualism occasionally argue that when we defer to standard usage, we do not understand ourselves as having had false beliefs or as having misapplied our own terms, but rather as having spoken a language ill-suited to our communicative interests. (See, for instance, Bilgrami 1992a, pp. 79-89, 110). However, as a *general* psychological claim, saying that deference is a response to such a norm relating explicitly to our desire to communicate has little plausibility.

²³ One should note that the problem cases for the individualist are not limited to those in which we are *actually* out of line with our community's usage. We may be disposed to defer to people who are not actually authoritative about the terms in our language (either because they have no real expertise, or that they are experts, but not in our own community).

of our concepts.²³ If the people to whom we are inclined to defer to are prevalent enough, then the sorts of alternate concepts that they bring are relevant alternatives even in those cases when no correction and deference occurs. Our mere disposition to defer when ‘corrected’ in a persuasive fashion is enough to undermine our self-knowledge. Just as the non-individualist’s problems with self-knowledge are not limited to those cases where we have actually switched concepts, the individualist’s problems are not limited to those cases where we actually defer to correction.²⁴

In addition to this, those who suggest that switching cases are prevalent enough to make them a relevant alternative even in those cases where no switching actually occurs may be overestimating how much switching the non-individualist is committed to. For instance, it has been suggested that the non-individualist is committed to one’s undergoing a ‘switch’ every time one spends time in (and is inclined to defer to) a new group that uses one’s words in a way different than how they had been used in one’s old group.²⁵ However, if my disposition to defer to the correction of an Englishman about my use of ‘chicory’ depends upon my being unaware that the English and American usages differ with respect to that term, then the disposition itself is not enough to commit the non-individualist to my concepts shifting. Things may be different if I actually engage in a significant amount of ‘chicory’ discussion in England without having any more in North America, but merely being in England will not be enough to alter the content of my chicory thoughts.²⁶ Switches may happen, but they will be less common than some might think.

A argument for prevalence of switching cases that is tied more closely to changes in one’s physical environment is provided by Keith Butler, who reasons as follows:

²⁴ Or for the conceptual role semanticists, those cases where we actually revise our beliefs.

²⁵ “Take, for example, a native speaker of American English, S, who travels to Britain and stays for an extended period of time; suppose she takes ‘chicory’ to mean the same in British English as it does in American English, though she has only incomplete knowledge of the meaning of the term; according to Burge and Putnam, she then defers to her language community, particularly the ‘experts,’ who fix the meaning of the term. As a matter of fact, however, ‘chicory’ means something different in American and British English. Because S’s language community has become relevantly different, S is a victim of slow switching.” (Butler 1997 p. 778. (For a more guarded description of such cases, see Ludlow 1995 pp.227-8).)

²⁶ Furthermore, if I continually ‘switch’ between the communities, it is likely that the meaning of the term will become indeterminate. Why this may happen, and why simply being in a new community that one is inclined to defer to is not enough to produce a content switch can be seen if the cases are understood in the framework suggested by Evans 1973, and discussed in more detail in Jackman 1996.

Assume (what I believe to be true) that there are subtle differences in constitution, too small really for non-expert cola drinkers to notice, between the colas sold by a certain brand X in Europe, and the colas sold by that same brand X in the U.S.; there are, in other words, E-colas and U-colas. Now imagine that S is an average cola drinker who has spent a fair amount of time in both Europe and the U.S., and has all the while been ignorant of the systematic differences between E-colas and U-colas. Externalism seems to entail that S will have developed, unwittingly, two concepts, one of E-colas and one of U-colas. But suppose that S thinks the thought she would express as “colas taste good”. What concept, according to externalism, is deployed in this thought? Regardless of which concept is in fact featured, it is clear that the other would be a relevant alternative. So, if S cannot tell the difference between E-cola-thoughts and U-cola-thoughts, S cannot know the contents of her thoughts; such is the dictate of externalism.²⁷

Butler goes on to suggest that similar arguments can be generated for beers, pizzas, flowers and any kinds in which “come in sub-varieties that bear superficial similarities to each other, but are nonetheless of different lower-order types.”²⁸ Butler’s argument, however, presupposes that, say, “Cola-X” should be treated as a natural kind term, when it more plausibly viewed as picking out a (for lack of a better word) “corporate kind.” Just as a rival companies ‘knock-off’ brand would not be Cola-X even if it had precisely the same formula, the European version of the Cola-X can still be of the same kind even if its internal make up is slightly different from the American version. What would produce a ‘switch’ would be if the “Cola-X” in Europe were produced by a completely different corporate entity that stumbled on the same name by coincidence. Such cases would, however, be far less frequent than the sort Butler imagines. A similar reply could be given for all of the purported cases of non-social switching that Butler presents, so the claim that the switching cases are prevalent for the non-individualist is still open to doubt.²⁹

The worry that unperceived changes are prevalent enough to make the alternate concepts relevant alternatives even when no change actually occurs is, then, if anything, *more* pressing if one is an individualist.

Knowledge of content may thus seem problematic whether one is an individualist or not.³⁰ Nevertheless, if we must learn to live with our lack of self-knowledge, it seems much easier to do so within a non-individualistic framework. After all, since the non-individualist allows that our contents are partially determined by facts about things ‘external’ to us, our occasional lack of

²⁷ Butler 1997, p. 779.

²⁸ Butler 1997, p. 779.

²⁹ Butler does point that non-individualists are rarely forthcoming about how one should pick out the relevant sorts of kind (p. 780), and I try to do so in Jackman 1996.

³⁰ Boghossian 1989 draws a similar conclusion, but for different reasons.

knowledge about our thought contents can be explained in terms of our ignorance of such facts.³¹ The non-individualist can explain the speaker's belief that his concept hasn't changed in terms of his ignorance of the fact that his physical or social environment has changed in a relevant way. However, it isn't clear that there is a similar factual mistake that the individualist can appeal to explain the speaker's mistake about his thought contents in deference cases. When Bert now claims that he had a false belief about arthritis, there is currently no factual mistake he can be said to make about what is going on around him. His purported mistake about the content of his earlier thought seems self-standing. While the non-individualist may have to admit that ignorance of fact can sometimes lead to mistakes about meaning and content, the individualist seems forced to say that we are sometimes *simply* mistaken about the content of our thoughts. The non-individualist can explain our mistakes in terms of mistakes about the world, the individualist must explain them in terms of mistakes about the nature of meaning and content.³²

In conclusion, the explanation of the intuition that there is a tension between non-individualism and self-knowledge relies on the assumption that knowledge of comparative content is needed for knowledge of content (at least whenever the 'compared' contents were relevant). However, if *this* is the source of the tension, then the individualist seems even worse off than the non-individualist. Considerations of self-knowledge, while often presented as one of the stronger considerations in favor of individualism, may actually provide us with a compelling reason to reject it.

³¹ Within such a framework it seems preferable to explain why our judgments are authoritative rather than how we know our contents.

³² The individualist might reply that his framework still allows that we always *can* have *a priori* knowledge of content, because we could, if we seriously reflected on such deference cases, come to 'correct' theory of meaning by *a priori* reasoning. *A priori* reflection could, in principle, lead Bert to a position similar to that of, say, Bilgrami's *Belief and Meaning*, and with it a correct conception of his earlier beliefs. Like Descartes, the individualist need not claim that the *a priori* knowledge in question is easy to get. However, this view is not only implausible, but seems to be pulling away from the sort of 'immediate' self-knowledge that it does seem intuitive that we have. Our intuitions about self-knowledge are tied to the intuition that we are typically authoritative about our thought contents, and this authority should not be dependent on our first going through a good deal of (individualistic) philosophical reflection on the nature of meaning.

Works Cited

- Bilgrami, A. 1992a, *Belief and Meaning*, Cambridge: Blackwell.
- Bilgrami, A. 1992b, "Can Externalism Be Reconciled with Self-Knowledge?" *Philosophical topics* 20: 223-267
- Boghossian, P. 1989, "Content and Self-Knowledge", reprinted in Ludlow and Martin 1998.
- Bueckner, A. 1990. "Skepticism and Knowledge of Content." *Mind* 99: 447-451.
- Bueckner, A. 1997. "Externalism and Memory", reprinted in Ludlow and Martin 1998
- Burge, T. 1979. "Individualism and the Mental", in French, Uehling & Wettstein, *Midwest Studies in Philosophy IV: Studies in Metaphysics*, Minneapolis, University of Minnesota Press, 1979.
- Burge, T. 1988, "Individualism and self knowledge", reprinted in Ludlow and Martin 1998
- Butler, K. 1997. "Externalism, Internalism, and Knowledge or Content", *Philosophy and Phenomenological Research* vol. LVII no. 4.
- Davidson, D. 1987, "Knowing One's Own Mind", reprinted in Ludlow and Martin 1998
- Davidson, D. 1989. "The Myth of the Subjective", in Krausz, M. (ed.). *Relativism: Interpretation and Confrontation*, Notre Dame: Notre Dame University Press, 1989.
- Davidson, D. 1994. "Donald Davidson", in Guttenplan, S. (ed.). *A Companion to the Philosophy of Mind*, Cambridge: Blackwell, 1994.
- Falvey, K. and Owens, J. 1994. "Externalism, Self-Knowledge, and Skepticism", *The Philosophical Review*, vol. 103, no. 1.
- Evans, G. 1973. "The Causal Theory of Names", in Evans, G. *Collected Papers*, New York: OUP, 1985.
- Fodor, J. and Lepore, E. 1992. *Holism: a shopper's guide*, Cambridge: Blackwell.
- Jackman, H. 1996. *Semantic Norms and Temporal Externalism*, Ph.D. Thesis, University of Pittsburgh.
- Jackman, H. 1998. "Individualism and Interpretation" *Southwest Philosophy Review*, V.14, n. 1, Jan 1998.
- Ludlow, P. 1995, "Externalism, Self-Knowledge and the Prevalence of Slow Switching", reprinted in Ludlow and Martin 1998
- Ludlow, P. 1996. "Externalism and Memory: A Problem?" reprinted in Ludlow and Martin 1998
- Ludlow and Martin (eds.), 1998, *Externalism and Self-Knowledge*, Stanford: CLSI Publications.
- Putnam, H. 1975. "The Meaning of 'Meaning'", in Putnam, H. *Mind Language and Reality*, New York: CUP, 1975.

Warfield, T. 1992, "Privileged Self-Knowledge and Externalism are Compatible", reprinted in Ludlow and Martin 1998.

Woodfield, A. 1982. "Introduction" to Woodfield, A. (ed.). *Thought and Object*, New York: OUP, 1882.