Social Externalism and First-Person Authority

Lynne Rudder Baker

University of Massachusetts Amherst

Abstract. Social Externalism is the thesis that many of our thoughts are individuated in part by the linguistic and social practices of the thinker's community. After defending Social Externalism and arguing for its broad application, I turn to the kind of defeasible first-person authority that we have over our own thoughts. Then, I present and refute an argument that uses first-person authority to disprove Social Externalism. Finally, I argue briefly that Social Externalism—far from being incompatible with first-person authority—provides a check on first-personal pronouncements and thus saves first-person authority from being simply a matter of social convention and from collapsing into the subjectivity of "what seems right is right."

Social Externalism and First-Person Authority

Externalism is the thesis that the contents of thoughts are often determined in part by various features of the thinker's environment. 'First-Person Authority,' as I am using the term, names the thesis that we typically have a priori knowledge of the contents of our own thoughts, where a priori knowledge is independent of justification by empirical evidence. There is a prima facie conflict between these two

theses: How can we know a priori what we think if the contents of our thoughts are partially determined by external features of our environment—features that can be justified only by empirical evidence? I shall try to answer this question. But first, I want to formulate and explore a specific thesis of externalism, and to explore the thesis of First-Person Authority. Then, I shall argue for the compatibility of the theses as I construe them. Since there are already acres of articles and books on these issues, all I can hope to do is to sow a new seed or two into this well-tilled field.

In 1. I shall set out my own construal of Social Externalism and argue for its broad application. In 2. I shall turn briefly to the kind of defeasible first-person authority that we have over our own thoughts. In 3. I shall present and refute two arguments that purport to show that Externalism and First-Person Authority are incompatible. Finally, in 4. I shall suggest that Social Externalism—far from being incompatible with first-person authority—provides a check on first-personal pronouncements and thus saves first-person authority from being simply a matter of social convention or from collapsing into the subjectivity of "what seems right is right."

1. Social Esternalism

Externalism is usually understood as a thesis about how contents of thoughts are individuated. There are two ways to understand this thesis—or rather two ways that environmental features may contribute to the content of a person's thought: The first interpretation focuses on the <u>thought</u>: What are the conditions under which a thought has the particular content that it has? The second interpretation focuses on the <u>thinker</u>: What are the conditions under which a person <u>has</u> a thought with a particular content? The significance of this distinction will emerge later, but the thesis that concerns me answers the second question: What are the conditions under which a person <u>has</u> a thought with a particular content? I shall call the thesis that concerns me 'Social Externalism' and formulate it as follows.

(SE) Most of the contents of a person's thoughts are determined (in part) by social and linguistic practices of the person's community.

The identity of a thought depends on the identity of its propositional content, canonically expressed by a 'that'-clause. To have a familiar framework, I'll use the word 'concepts' to refer to the items constitutive of propositional contents, and I'll use the word 'thoughts' for any mental states with propositional content—e.g., beliefs, desires, intentions, hopes, fears, etc. I'll say that thoughts

contain or are made up of concepts.   E.g., the thought that snow is white contains the concept *snow* and the concept *white*.[1] I do not intend for my use of 'thought' or 'concept' to carry much theoretical weight.   (For simplicity, I'll put aside singular thoughts that contain individuals, if there are any.)[2]   So, concepts make up thoughts.

What is significant about concepts (besides the fact that they make up thoughts) is that they have conditions of application: a concept definitely applies to some things and definitely does not apply to other things. Many concepts (e.g., honesty, harrassment) do not have such clear-cut application conditions, and so whether or not one has mastered such concepts is not a clear-cut matter.   Moreover, one's knowledge of the application conditions of a concept may be only partial or incomplete, as Tyler Burge has forcefully taught us. (Burge, 1979) Indeed, we have many thoughts made up of concepts of which we have only limited understanding: you may believe that scientists talk about black holes, or that Beethoven wrote nine symphonies without complete mastery of the concept *black hole* or the concept *symphony*.

Since complete mastery of the application conditions of a concept is not required in order to have thoughts containing that concept, we need a term to signify the limited kind of mastery that is required to have a concept.   I'll use the weak word 'has' in order to

have a term that has not already been appropriated by the literature. Letting 'S' range over thinkers, say:

> S has a concept C if and only if (i) S is able to apply C correctly in a significant range of cases; and (ii) S has some (perhaps partial) understanding of C's conditions of application.

Of course, the extent of the significant range of correct application of C and the degree of understanding required are both vague, but I'll try to use examples that are not controversial. Now let me introduce a name for the concepts contained in thoughts to which Social Externalism applies. Call them 'SE-concepts'. Then:

> Concept C is an SE-concept if and only if: For any S, if S has C, then holding constant S's physical history and physical internal states, if S had been in a community with relevantly different social or linguistic practices, S would have failed to have C.

where a person's physical history includes all the nonintentional interactions that a person has with physical entities in the environment, and physical internal states are nonintentional states (like brain states).[3]   So, a concept C is an SE-concept just in case:

whether or not someone has the concept C depends on the social or linguistic practices of her community.

An SE-thought is a thought containing an SE-concept. A thought is an SE-thought if and only if it contains an SE-concept, and a concept is an SE-concept if and only if whether or not someone has it depends on social and linguistic practices. Then: Social Externalism is the thesis that many of our thoughts are SE-thoughts. Most people, I suspect, are prepared to regard social, political, and legal concepts as SE-concepts. However, the scope of Social Externalism is far broader than just the thoughts containing the obviously SE-concepts.

Sometimes, it is erroneously assumed that Social Externalism applies only to thoughts involving concepts that the thinker partially misunderstands. Burge made clear that "communal practice is a factor...in fixing the contents of my attitudes, even in cases where I fully understand the content." (Burge, 1979, p. 85.) Or again:

> [E]ven those propositional attitudes not infected by incomplete understanding depend for their content on social factors that are independent of the individual, asocially and non-intentionally described. For if the social environment had been appropriately different, the contents of those attitudes would have been different."

Partial understanding is a sufficient but not a necessary condition for thoughts to which Social Externalism applies. Let me illustrate with a case of someone with full understanding of a concept that is not intuitively a social concept, but which turns out to be an SE-concept: Suppose that Sam believes that boulders are heavy and that Sam expresses his belief in English by saying, "Boulders are heavy." To see that 'boulder' expresses an SE-concept, consider an alternative linguistic community that has a word that sounds like 'boulder' that includes only large rocks found on dry land (in fields or on mountains), and not to large rocks found in or near bodies of water, say.

Then, Sam, who makes no mistake about the concept *boulder* in our community, and believes that boulders are heavy would not have that belief in the alternative linguistic community that uses what sounds like 'boulder' to refer only to land boulders. If Sam had been brought up in the alternative community and uttered what sounds like 'Boulders are heavy,' he would not thereby have expressed a belief that boulders are heavy, but rather a belief that we should translate without using our word 'boulders'. The concept that we express by 'boulder' in our community has different application conditions from the concept that is expressed by what sounds like 'boulder' in the alternative community. The concepts so expressed thus differ in the two communities. Sam—without any "internal" difference or

difference in physical history—would have different beliefs in the different linguistic communities.   Such beliefs are thoughts to which Social Externalism is applicable. The concept *boulder*, though not an intuitively social concept, is nevertheless as SE-concept.

Even natural-kind concepts turn out to be SE-concepts. Limestone is calcium carbonate ($CaCO_3$); sandstone is silicon dioxide (SiO_2).[4]   Limestone is usually buff-colored, and sandstone, though usually reddish-brown, can be buff-colored as well.   Limestone and sandstone are composed of different kinds of molecules, and are of different natural kinds. I want to argue that the concept *limestone* and the concept *sandstone* are SE-concepts, despite the fact that they are natural-kind concepts.

Consider someone, call him 'Joe', who lives in a community where there is a lot of limestone and no sandstone.   Joe speaks the local language and knows a lot about limestone. He knows that limestone has a buff color and is sometimes used as a building material. Joe has never seen any limestone, nor been in physical contact in any way with limestone. But he has learned about limestone from books and classmates in school. Joe has the concept *limestone* in the sense specified: (i) Joe is able to apply the concept *limestone* correctly in a significant range of cases (indeed if he were to come across some limestone, he might well infer that it was

limestone); and (ii) Joe has some (perhaps partial) understanding of the application conditions of *limestone.* So, Joe has the concept *limestone*.

Now suppose that Joe had grown up in an alternative community, where there was no limestone, but a lot of buff-colored sandstone. In the alternative community, Joe speaks the local language and knows a lot about sandstone. He knows that sandstone has a buff color and is sometimes used as a building material. Joe has never seen any sandstone, nor been in physical contact in any way with sandstone. But he has learned about sandstone from books and classmates in school. In the alternative community, suppose that Joe had exactly the same physical interactions with exactly the same people and other things that he had in the actual community. Whatever the word was for limestone in the actual community had a homonym used in the same way to designate the sandstone in the alternative community that had no limestone. If Joe had been brought up in the alternative community, he would not have had the concept *limestone* at all.   Rather, he would have had a different concept— *sandstone*. But in the alternative community, Joe would refer to sandstone by what sounded like the same word that, in the original community, he used to refer to limestone.

The concept *sandstone* is a different concept from the concept *limestone*. The two concepts have entirely different application conditions. Now, varying only Joe's community, we can vary which concept he has--*limestone* or *sandstone*--without changing any of his internal states or any of the people or things that he had interactions with or any of the nonintentional ways that he interacted with them. He did not even have any false beliefs about the application conditions of either of the concepts *limestone* or *sandstone*; he had no false beliefs about the chemical composition of either since he had no beliefs about chemical composition—any more than our ancestors who had the concept *limestone* did.

So, *limestone* is a natural-kind concept that satisfies the definition of an SE-concept. Granted, in addition to the different social or linguistic practices of the two communities, there is a physical difference as well (there is no limestone in the alternative community). But this is irrelevant to whether or not *limestone* is an SE-concept. An SE-concept is one such that whether someone has it or not depends on the social and linguistic practices of the community.

Let me entertain an objection: "What the limestone/sandstone story shows is only that natural-kind externalism holds; the difference betweeen Joe's having the concept *limestone* as opposed to *sandstone* does not depend on linguistic practices at all," the

objection goes. "It only depends on the difference of natural-kinds. Social and linguistic practice is irrelevant to which concept Joe has."

I have two responses: (1) In the story as told, Joe would not have either concept *limestone* or *sandstone* except for the social and linguistic practices of the two communities. He has never had any interaction with either limestone or sandstone; indeed, the existence of samples of a natural kind is <u>never</u> enough for someone to have the concept of the kind. Social and linguistic practice is the required intermediary for having shared concepts at all.[5] So, social and linguistic practice is not irrelevant to which concept Joe has.

(2) Moreover, and this is my second response, for purposes of understanding the acquisition of concepts, practices should be individuated broadly, so that what counts as a social or linguistic practice includes features of the environment that get referred to. This is so, because concepts are individuated by their application conditions. A concept that applies to Fs is distinct from a concept that applies, not to Fs, but to Gs. So, practices that underwrite the concept *limestone* concern calcium carbonate, and are distinct from practices concerning silicon dioxide. Again, the mere existence of limestone in a community is never sufficient for someone to have the concept *limestone*.

Of course, the '*limestone*' example is a slight variation on Burge's old story. I tell it again to emphasize that it is not just intuitively social concepts like *speed limit* or *driver's license* that are SE-concepts, but also natural-kind concepts like *water* or *limestone* are SE-concepts as well. Given the way that I have specified what I mean by 'SE-concept,' what makes a concept an SE-concept is not a matter of what individuates the concept; it is a matter of what is necessary for someone to <u>have</u> the concept. The fact that a concept is <u>individuated</u> independently of linguistic or social practices (as is the concept *limestone*) is irrelevant to whether or not it is an SE-concept. A concept is an SE-concept if one's <u>having</u> it depends on linguistic or social practices, regardless of how the concept is individuated . (If complete mastery of individuating conditions of a concept were required to have a concept, most of us would be woefully short of concepts.)[6] So, Social Externalism, as I construe it, applies to concepts whose possession depends on social or linguistic practices. Now we can see why I distinguished at the outset between two ways of understanding social externalism: The thesis that applies to the <u>having</u> of concepts is the one that interests me (and, I think, Burge).

In short, the range of Social Externalism is vast. Social Externalism applies, not only to thoughts containing concepts that are intuitively social concepts, but to thoughts containing natural-kind

concepts (like *limestone*) whose individuation is independent of social or linguistic practices. Whether social or linguistic practices contribute to individuating a concept is irrelevant to whether social or linguistic practices contribute to someone's having the concept. And one's thoughts are limited by the concepts that one has.

2. First-Person Authority

As I said at the outset, the thesis that I am calling 'First-Person Authority' is this:

(FPA) A person knows a priori the contents of her own thoughts,

where a priori knowledge requires no justification by empirical evidence. (FPA) is woefully sketchy as an account of first-person authority. Our knowledge of our own thoughts is immediate, noninferential, non-observational; it is first-personal knowledge "from the inside," so to speak. We know what we think in a way differently from the way we know anything else. Our first-person authority is rooted in our being rational agents.[7] Any full account of first-person authority will have to go into these vexing matters. Fortunately, all that is needed for purposes here is the characterization of first-person authority as a priori (nonempirical) knowledge of the contents of our

own thoughts.   So, let me just make a few comments about how I intend (FPA) to be understood.

First and most important, claims to first-person authority are not infallible. We typically know what we believe, fear, hope, desire and intend; but we can be (and routinely are) corrected by others—friends, therapists, confessors. So, pace Descartes, claims to first-person authority are defeasible. Our beliefs about our thoughts are subject to correction. The more morally and personally significant our thoughts, the more likely that our claims to know them are mistaken. E.g., we are more likely to be wrong about our beliefs about our own honesty than about our beliefs about water.

Most prominently, our claims to first-person authority are defeated by self-deception. E.g., one may sincerely and comprehendingly say, "I fear that my work will be interrupted this afternoon" when in fact she hopes that her work will be interrupted this afternoon. And there is ample empirical evidence that people are unreliable in their reports of their own reasoning processes. (Nisbett and Wilson, 1997; Nisbett and Ross 1980)   On the one hand, First-Person Authority is the "default position" in mundane cases—like our knowing the contents of our standing attitudes and of thoughts of which we are currently conscious. To doubt someone's first-person pronouncements about her own thoughts, one needs some special

reason. On the other hand, there are numerous defeating conditions for claims to first-person authority.

So, first-person authority is extensive, but claims to first-person authority are <u>defeasible</u>.   Indeed, it is defeasibility that promotes our beliefs about what we think to knowledge. The fact that we can make errors, and that our errors can be corrected by others, gives legitimacy to our claims to know our own thoughts.

Second, in order to have First-Person Authority over one's thoughts, one must meet several conditions:

(i) One must have conscious thoughts—thoughts that "can occupy the attention of the thinker." (Peacocke, 1992, p. 152) over which to exercise first-person authority.

(ii) One must have a first-person perspective in order to recognize thoughts as her own. (In Baker, 2000, I argued that first-person perspectives are what distinguish persons from all other kinds of beings.) First-personal knowledge of my own thoughts requires that I be able to recognize my thoughts as my own.   I must be able to attribute them to myself.

(iii) One must have concepts of relevant propositional attitudes.   In order to <u>know</u> that one believes, desires, intends, hopes, or fears something, one must have concepts of believing, desiring,

intending, hoping or fearing. So, it is not the case that one must have a concept of belief in order to have any beliefs (pace Davidson (1975)),[8] but one does have to have a concept of belief in order to have first-person authority over one's beliefs.

If I were to develop a full account of first-person authority, I would want to avoid both a perceptual model that takes us to be spectators of our own minds, and a deflationary account that denies that we have a priori knowledge of our thoughts. (Boghossian, 1989).[9] In other areas, we can know or make correct judgments about matters of fact without any justification by observation. After all, we know the positions of the limbs of our body without observation (proprioception).[10] And, as Wittgenstein points out, some of our judgments about what time it is—"It must be nearly noon by now"— are based on no observation (Wittgenstein (1953) ¶ 607; cited by Moran, 2001).   These cases, as well as the cases of our knowledge of our own minds, concern things that we can be right or wrong about and hence are subject to epistemic evaluation. I shall return to this point later, but now let's turn to the question of the compatibility of first-person authority and social externalism.

3. Arguments Against Social Externalism and First-Person Authority

At the beginning, I raised the question: How can we know a priori what we think if the contents of our thoughts are determined by external features of our environment—features that can be only known empirically? Now I want to answer the question with respect to the two theses, as I have construed them—Social Externalism and First-Person Authority.   Here again are the two theses:

(SE) Most of the contents of a person's thoughts are
    determined (in part) by social and linguistic practices of
    the person's community.

(FPA) A person knows a priori the contents of her own thoughts.

I'll consider and reject two arguments against the compatibility of Social Externalism and First-Person Authority.[11]

The first argument is that Social Externalism, together with First-Person Authority, leads to the implausible conclusion that we have a priori knowledge of empirical facts. The thinking goes like this: If Social Externalism is true, then the contents of our thoughts are determined by social and linguistic practices; if there are social and linguistic practices, then there exist other people. The facts that there are social and linguistic practices and that there exist other

people are empirical facts that follow from Social Externalism. So, it seems to follow from Social Externalism that if we know a priori the contents of our own thoughts, we must know a priori that there are social and linguistic practices and other people.

Burge dismisses this line of thought, by arguing that one may know something without knowing the "background enabling conditions" that make that knowledge possible. For example, your knowledge of your thought that water is a liquid does not require knowledge of the conditions that make the thought possible—e.g., the existence of H_2O. (Burge, 1988, p. 118). Although I think that Burge is right about this, I want to take another tack. Consider Argument A —an illustration of the above line of thinking against the conjunction of First-Person Authority and Social Externalism:

Argument A

(1) I know a priori that I believe that boulders are heavy. (First-Person Authority)

(2) I know a priori that: If I believe that boulders are heavy, then there exist other people. (Putatively Deducible from Social Externalism)

So, (3) I know a priori that there exist other people.

I shall assume that being known a priori can be transferred from (p &
p→q) to q, and hence that Argument A is valid. So, the question is
whether the premises are true. The first premise is an instance of the
thesis of First-Person Authority and hence is acceptable. According to
the second premise, what is alleged to be known a priori is a
conditional that is supposed to follow from Social Externalism:

(H) If I believe that boulders are heavy, then there exist other
people.

Granted, if Social Externalism is true, then (H) is true. But do I know
(H) a priori? A priori knowledge of (H) is supposed to follow from a
priori knowledge of Social Externalism. In that case, I know a priori
that (H) is true only if I can justify Social Externalism a priori. Is my
justification for Social Externalism a priori? I don't think so.

The justification for Social Externalism rests on thought
experiments of the sort that I have presented. As Timothy Williamson
has argued in his 2004 Presidential Address to the Aristotelian
Society, the use of thought experiments need not involve any a priori
intuition, but only "our general cognitive capacity to handle
counterfactual conditionals, which is not exclusively a priori...."
(Williamson, 2004, p. 1)[12]   That is, Williamson assimilates the crucial
parts of thought experiments in philosophy to ordinary, contingent
counterfactuals that are not asserted on a priori grounds. E.g., "'If Dr.

Johnson had kicked George Berkeley, he would have kicked a bishop'

is both a posteriori and contingent." (Williamson, 2004, p. 10)[13]

We have a general cognitive ability to assess the truth value of

such everyday counterfactuals, and there is no reason to believe that

the evaluation of philosophical counterfactuals requires any different

cognitive capacities. (Williamson, 2004, p. 13)   Evaluation of

counterfactuals requires a mass of background beliefs, which include

many empirical beliefs that cannot be explicitly specified. To evaluate

a counterfactual, the background knowledge needed is "a general

sense of how things go, honed over long experience." (Williamson, p.

6) Thought experiments do not involve any sui generis philosophical

thinking, but only the application of our general capacity to process

counterfactuals, a capacity "widely used throughout our cognitive

engagement with the spatio-temporal world." (Williamson, pp. 20-21)

So, if Williamson is right (as I think he is), justification of Social

Externalism by thought experiments does not license premise (2).[14]

To sum up the response to Argument A: Premise (2) is true only

if the justification of Social Externalism provided by the thought

experiments is a priori. But thought experiments should be

understood in terms of counterfactuals, the evaluation of which does

not involve "fundamentally different cognitive capacities from

evaluations of counterfactuals that we classify as a posteriori."

(Williamson, p. 14) Since our justification for Social Externalism is not a priori, we cannot know (H) a priori. Hence, premise (2) is false. It is not the case that I know a priori that: If I believe that boulders are heavy, then there exist other people. Social Externalism can be the source of empirical conclusions, but not the source of empirical conclusions known a priori.

Now turn to the other claim that threatens the compatibility of Social Externalism and First-Person Authority. Assuming that we avoid the inference from Social Externalism to a priori knowledge of empirical facts, there is another threat: If Social Externalism is true, then it seems that the thesis of First-Person Authority is false. Now here is a simple argument against the compatibility of Social Externalism and First-Person Authority:

Argument B

(4) If Social Externalism is true, then we do not know the
     contents of our own thoughts.[15]

(5) If we do not know the contents of our own thoughts, then the
     thesis of First-Person Authority is false.

∴ (6) If Social Externalism is true, then the thesis of First-Person
     Authority is false.

The anti-externalist has an argument for premise (4)—cases based on "slow-switching". (Boghossian, 1989)   Here is such an anti-externalist argument for (4): Suppose that Beth travels between two linguistic communities—ours, in which the word 'red' expresses the concept *red*, and another, in which there is a word that sounds like 'red', but it expresses the concept that we would call *red-or-orange*. So, we should translate the word in the other community (the word that sounds like 'red') not as 'red' but as something else, say, 'ored'. There is no single English word that is a synonym for what sounds like 'red' in the other community. Now Beth lives in Australia for many years, then moves to the other community. She does not notice the difference in the linguistic practices of the two communities; as it happens, she never hears oranges called (what sounds like) 'red'.

When Beth first arrives in the other community and says 'Roses are red', she expresses her belief that roses are red.   But, continues the anti-externalist, Social Externalism implies that after living in the other community for decades, when Beth says, 'Roses are red,' she no longer expresses her original belief that roses are red. For the concept expressed by the word 'red' in the other community is not *red*, but *red-or-orange*, as we would put it. By assimilating the other community's language, Beth's belief changed without her realizing it.[16] So, the anti-externalist charges, according to Social Externalism, now when she says what sounds like 'Roses are red', she

—sincerely, but unwittingly—expresses a different belief (viz., that roses are ored).   Since she would not be able to tell the difference between her original belief and her new belief, she would not know the content of her own belief about roses. Hence, the anti-externalist says, Beth does not know the contents of her own thoughts.

Burge, I think, would reject the example of Beth, and reject premise (4). After living in the alternative community for decades, Beth does come to speak the other language. When she thinks the simple thought that she would express by what sounds like "Roses are red", the thought that she thinks is, indeed, that roses are ored. But, according to Burge, she knows that she is thinking the thought that roses are ored.   Her knowledge that she is thinking that thought "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." (Burge, 1988, p. 118) Her thoughts about the color of roses contain the concept that her community uses, a concept that we translate as '*ored*'.[17] So, years after the switch, what she knows when she knows the contents of her own mind is that she believes that roses are ored. Hence, on this Burgean view, she does know the contents of her own mind, and the thesis of First-Person Authority is not threatened by Social Externalism.

Even if Burge is correct here, Beth is still ignorant of the fact that she has undergone a change of belief. She falsely believes that she still has the same belief that she expressed 25 years ago by saying what sounds like "Roses are red." And she cannot discover that this belief is false without recourse to empirical information. However, her ignorance of the difference in the two concepts expressed by what sounds like 'red' does not threaten the thesis of First-Person Authority. Beth knows that she has a belief that her belief about the color of roses is unchanged; what she doesn't know is that her belief that her belief is unchanged is false. But not knowing that a belief about a belief is false is irrelevant to the thesis of First-Person Authority.

In order for Beth's case to be a challenge to First-Person Authority, the anti-externalist needs to add a condition on knowledge, such as the following. Call it 'the Discrimination Principle':

> (DP) A person knows that p only if she can distinguish her situation from relevant alternative situations in which p is false.

If we accept (DP) and take the near-duplicate cases to be relevant alternatives, then we should say that Beth knows that she believes that roses are ored only if she can distinguish her situation from relevant situations in which it is false that she believes that roses are

ored.[18] But since Beth cannot distinguish her "ored" situation from her former "red" situation, anti-externalists who accept (DP) may deny that she knows the contents of her own thoughts.

But there is reason not to accept appeal to (DP) in such cases where the situations are exactly the same except for a difference in social or linguistic practices: The Discrimination Principle would rule out too much of what we reasonably think that we know. Think of concepts like *disability*, *sexual harrassment*, *professor*.   For many, if not most, of the concepts that I have, for example, you could make up near-duplicate cases that I could not distinguish from the actual world, but in which there were closely related concepts, but not the concepts that I have. In that case, if we accepted (DP) as a constraint on what can be known, and took near-duplicate cases to be "relevant alternative situations," I could not have any thoughts containing those concepts. An anti-externalist may bite the skeptical bullet and deny that we know any propositions containing any such concepts. That is surely the wrong response: Concepts like *disability*, *sexual harrassment*, *professor* are social on their face and hence do depend on community practices.   And I am confident that I have thoughts containing all of them and that I know the contents of those thoughts. But if (DP) were a plausible constraint, I would not have thoughts containing any of those concepts. Thus, we have good reason to rule out these near-duplicate cases as relevant

alternatives in the context of the Discrimination Principle, or to reject the Discrimination Principle altogether.

So, dismissing the relevance of the Discrimination Principle to Beth, we can conclude that Beth knows the contents of her own thought that she expresses by saying what sounds like "Roses are red," and the fact that she is unaware of her change of belief is no threat to First-Person Authority.

To sum up the response to Argument B: The argument in favor of premise (4) falls short. The case of Beth provides no reason to think that (4) is true. (And I doubt that any better argument for (4) will be forthcoming.) Hence, I see no reason to doubt that the thesis of First-Person Authority is compatible with Social Externalism. Indeed, I now want to suggest that, far from clashing with First-Person Authority, Social Externalism supports it.

4. Social Externalism in Support of First-Person Authority

Social Externalism allows for congruence between what thoughts you attribute to me and what thoughts I believe that I have. If I tell you that I believe that boulders are heavy, then you can know what I believe. The content of my belief—that boulders are heavy— has a public meaning. Social Externalism supports the claim that our thoughts are intersubjectively knowable, that what I think to myself

can be known by you.   The first benefit of Social Externalism is that it makes a place for first-person authority without solipsism.

Moreover, Social Externalism provides the basis for an intersubjective standard of correctness. When I claim to know that I believe something, my claim can be evaluated (and on occasion overruled) by a suitably-situated observer.   In the absence of a standard of correctness, it is difficult to see how there could be knowledge of our own thoughts at all.   Social Externalism underwrites the intersubjectivity of SE-thoughts, and the intersubjectivity provides a check on what one takes to be self-knowledge. Social Externalism lays the ground for the defeasibility of First-Person Authority.

My beliefs about my own thoughts are individuated by the same 'that'-clauses that your attributions to me are.[19] As Barry C. Smith put it, "[P]ublicity ensures that self-knowledge lays claim to a genuine subject-matter, the external dimension ensuring that whether a subject is in a given mental state is an objective, or at any rate, intersubjective, matter answerable to more than just the subject's opinions." (Smith 1998, p. 395-6) Social externalism thus brings together first-person and third-person points of view without collapsing either into the other.[20]

Some philosophers take the hallmark of mental states sto be the absence of any distinction between appearance and reality about the contents of one's mind.[21] I think that such a view is badly mistaken. Not only are self-deceptive cases counterexamples to the claim, but also the claim has disastrous consequences for the reality of mental states. If there is no difference between appearance and reality, then, as Wittgenstein said, "One would like to say: whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'." (Wittgenstein, 1953, ¶ 258)   Social Externalism saves first-person authority from collapsing into the subjectivity of "what seems right is right."[22]

References

Baker, L. R. (2000). Persons and bodies: a constitution view. (Cambridge: Cambridge University Press)

Boghossian, P. (1989). Content and self-knowledge. Philosophical Topics, 17, 5-26

Brown, J. (2004). Anti-individualism and knowledge, Cambridge, MA: MIT Press)

Burge, T. (1979). Individualism and the mental. (In P.A. French, T.E. Uehling, Jr., H. K. Wettstein (Eds.), Studies in metaphysics, Midwest

studies in philosophy IV (pp. 73-121). Minneapolis: University of Minnesota Press)

Burge, T. (1988). Individualism and selff-knowledge. Journal of Philosophy, 85, 649-663 (Reprinted in P. Ludlow and N. Martin (Eds.), Externalism and self-knowledge (pp. 111-127). Stanford: CSLI Publications)

Davidson, D. (1975). Thought and talk. (In S. Guttenplan (Ed.), Mind and Language. (pp. 7-23). Oxford: Clarendon Press)

Dretske, F. (1970). Epistemic operators. Journal of Philosophy, 67, pp. 1007-1023

Goldman A. (1986). Epistemology and cognition. (Cambridge, MA: Harvard University Press)

Lewis, D. (1996). Elusive knowledge. Australasian Journal of Philosophy, 74, 549-567.

Moran, R. (2001). Authority and estrangement: an essay on self-knowledge. (Princeton: Princeton University Press)

Nisbett, R.E. and Wilson, T. (1977). Telling more than we can know: verbal reports on mental processes. Psychological Review, 84, 231-259

Nisbett, R.E. and Ross L. (1980). Human Inference: Strategies and Shortcomings of Social Judgment. (London: Prentice Hall)

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

Smith, B. C. (1998). On knowing one's own language. (In C. Wright, B.C. Smith, and C. MacDonald (Eds.), Knowing our own minds (pp. 391-428). Oxford: Clarendon Press)

Williamson, T. (2004). Armchair philosophy, metaphysical modality and counterfactual thinking. Proceedings of the Aristotelian Society, 105, 1-23.

Wittgenstein, L. (1953). Philosophical investigations, ¶ 607. G.E.M. Anscombe, (Trans.) (New York: Macmillan)

[1] Concepts are distinct if they differ in their potential informativeness. For further discussion, see Peacocke (1992).

[2] Even if the content of a singular term is determined by the thinker's causal relations, and is independent of social or linguistic features, the thought containing the singular term may well be an SE-thought. E.g., the thought expressed by 'That is a dog' contains the SE-concept *dog*. See below.

[3] Say that x (an entity, property, thought, anything) is intentional if and only if either x is a propositional attitude or the existence or occurrence of x implies that there exist beings with propositional attitudes.

[4] According to the U.S. Geological Survey, both limestone and silicon are sedimentary (i.e., they have layers) and both are cementatious (i.e., they are formed by tremendous pressure, but not enough to produce crystallization). Marble, like limestone, is also made of calcium carbonate; but it has undergone metamorphic recrystallization.

[5] The fact that natural-kind terms are rigid designators is a matter of linguistic practice. A term's reference depends on how it is used.

[6] Indeed, I think that Wittgenstein's discussion of rule-following has given us reason to be skittish about the idea of the totality of application conditions of a concept.

[7] See See Richard Moran, *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton: Princeton University Press, 2001)., for development of this idea.

[8] Although non-language users (e.g., nonhuman animals) may have beliefs and other thoughts, the range of thoughts that they can have is severely circumscribed.

[9] Moran (2001, p. 16) points out that Boghossian's argument for the conclusion (that if externalism is correct, then we can't have introspective knowledge of our own thoughts) presupposes a perceptual model of introspection.

[10] For fuller discussion of knowing one's "own basic movements and bodily position without having to observe anything, internally or externally, see Moran, 2001, pp. 19-20.

[11] For a comprehensive treatment of the arguments for the (in)compatibility of Externalism and First-Person Authority, see Brown (2004).

[12] I am grateful to Hilary Kornblith for bringing this address to my attention.

[13] Note that the relation between antecedent and consequent is not causal.

[14] Williamson's point is that any distinction between the a priori and the a posteriori "lacks the significance with which it is often credited. It does not cut at the joints." He is not claiming that we can make no distinction between the a priori and the a posteriori, only that "such a distinction lacks the significance with which it is often credited." (p. 14)

[15] That is, all our undefeated thoughts—where a thought is undefeated if there is no nonarbitrary condition that exempts it from being known a priori.

[16] The internalist who takes thoughts to contain internal concepts that a thinker can have independently of social and linguistic factors denies that change of language results in change of belief. The externalist holds that our only access to the contents of our thoughts is by the sentences expressing them; so, thoughts had better match up with sentences expressing them, on pain of incoherence. The point at issue here, however, is not whether the internalist or externalist is correct overall, but whether Social Externalism (which allows that change of language may result in change of belief) precludes First-Person Authority.

[17] What is being translated by '*ored*' is the name of the concept. The term '*ored*' mentions the concept translated; the term does not use the concept.

[18] This additional condition—though unsatisfactory, as we shall see—is not ad hoc. Philosophers like Dretske (1970), Goldman (1986) and Lewis (1996) have proposed similar conditions on knowledge generally.

[19] This is a logical or semantic claim, not a psychological or neurophysiological claim. This claim is grounded in the need for coherence, not in any empirical theory.

[20] This is a major conclusion of Richard Moran in *Authority and Estrangement.*