Internalism, Externalism, and Accessibilism

(working title)

Brie Gertler, University of Virginia

Conee and Feldman (2001) observed that the term “internalism”, as used in epistemology, is ambiguous. It sometimes denotes the view that justification supervenes on factors within the thinker: specifically, on her mental states and properties (which are assumed to be “internal” to her). But at other times it refers to the view that justification is accessible to the thinker. Conee and Feldman labeled these views “mentalism” and “accessibilism”, respectively.

As used in the debate about mental content, “internalism” closely corresponds to mentalism. Content internalism is the thesis that thought content—specifically, cognitive value, the aspect of content that figures in rational explanations of psychological processes—supervenes on factors internal to the thinker. Content externalism is the denial of this thesis.1 Strikingly, however, arguments on both sides of this debate are focused largely on the question of accessibility about content. Internalists allege that externalism is incompatible with our access to our thought contents: in particular, with our ability to genuinely comprehend (and not just blindly track) what we’re thinking, through introspective reflection. Externalists charge that internalism overestimates our access to our thoughts, by implying that we can explicate them (i.e., uncover their truth conditions) through mere reflection.

In this paper, I argue that the focus on thinkers’ access to their thoughts is misplaced. Claims about reflective or explicatory knowledge of thought contents, which concern (higher-order) thoughts about thoughts, will not settle the debate between externalism and internalism. This debate turns, instead, on the significance of external factors for rational relations between first-order thoughts. Shifting the focus of the debate away from issues of accessibility neutralizes the leading, “incompatibility” objections to externalism. But this shift also neutralizes the key considerations marshalled in support of externalism. So my argument ultimately supports internalism.

The paper is divided into two parts. Part 1 argues that, while externalism may be incompatible with the claim that thinkers generally have access to the truth conditions of their thoughts, this incompatibility does not threaten externalism. For this accessibility claim is dubious. The most promising argument in its favor employs the premise that introspective access to truth conditions is a requirement of rational thought. But, I argue,

---

1 I argue elsewhere (Gertler 2012) that “content internalism” and “content externalism” are irremediably ambiguous, in that no way of specifying the boundary of the thinker—and thereby cashing out the crucial idea of what is internal to the thinker—will fit with the standard ways of classifying particular views as externalist and internalist, and standard views about the commitments of externalism and internalism. But I put that reservation aside for present purposes.
a rational thinker need not have reflective access to truths about her thoughts. Rationality is primarily a matter of relations among first-order mental states.

The second part of the paper addresses the real question at issue in this debate: whether external factors directly affect rational relations between first-order thoughts. I argue against this externalist claim by developing and applying a qualified version of Frege’s “differential dubitability” test for sameness and difference of cognitive value. The leading externalist response to this kind of Fregean argument interprets it as resting on an implausible accessibility thesis. I show that this interpretation confuses internalism with accessibilism. The Fregean argument for internalism is exclusively concerned with relations between first-order states, and does not rest on an accessibility thesis. And internalism is neutral about whether thoughts’ truth conditions are reflectively accessible to the thinker. 2 For clarity and ease of exposition I will focus on Burge’s externalist arguments. But the paper’s conclusions apply to the broader debate.

Part 1: Externalism and access

1.1 Externalism’s alleged incompatibility with introspective access

The alleged incompatibility between externalism and introspective access has been discussed at great length. There are two basic versions of the incompatibility objection.

The first version charges that externalism is incompatible with the fact that we can know what we’re thinking through introspection, or that introspective judgments are certain or especially reliable. The externalist may of course reject one or more of these claims about introspective self-knowledge. But he need not do so, since externalism does not appear to face any special problems here. (Wikforss (2008) makes this case in an especially persuasive and illuminating way.) These claims pertain to how we know our thoughts or attitudes: how higher-order judgments like “I am now thinking that \( p \)” or “I believe that \( q \)” relate to the lower-order thoughts they concern. The outlook on those issues is not constrained by the supposition that some thought contents are individuated by reference to external factors. Perhaps most crucially, the externalist can simply maintain that higher-order judgments inherit their contents from the lower-order thoughts they concern (Burge 1988, Gibbons 1996). This inheritance of content may occur through introspection; and it may mean that these higher-order judgments are especially likely to be true. 3 So externalism withstands the first version of the incompatibility objection.

---

2 Wikforss (2008) argues for a similar conclusion: in her words, “the real challenge posed to externalism concerns not self-knowledge, but understanding or concept grasp.” I encountered her argument as I was completing this paper. While our arguments are similar in spirit, they differ in some details and, especially, in how we develop the case for internalism. The addition of my arguments strengthens her already strong case for the conclusion that the shift away from questions of explication favors the internalist.

3 The inheritance of content might occur through the kind of containment relation posited by Burge (1988), or it may occur some other way: e.g., through an appropriate causal connection between lower-order and higher-order thoughts. So the content inheritance strategy can be used in conjunction with a variety of epistemic views about self-knowledge.
The second version of the incompatibility objection poses a greater challenge. It is motivated by the idea that our access to what we are thinking goes beyond the ability to reliably track our thoughts, or to form higher-order beliefs that inherit their contents. For possessing those abilities does not ensure that the thinker truly understands her current thoughts. As Goldberg says, “one can count as knowing that one is thinking that \( p \) without knowing what one is thinking in (knowing that one is) thinking that \( p \).” (Goldberg 2003, 254) The second version of the incompatibility objection alleges that externalism is incompatible with our ability to genuinely comprehend what it is that we’re thinking.

The claim that thinkers can comprehend their thoughts is open to various interpretations. The interpretation that fuels the current objection seems to be this. Through introspective reflection, rational thinkers can achieve a meaningful grasp of the contents of their thoughts—a grasp that enables them not just to think thoughts that inherit those contents but to explicate them, that is, to identify (substantive) truth conditions for them. Following Burge, I will use the term “explicatory knowledge” to refer to knowledge of the truth conditions of one’s thoughts (and/or of the criteria for satisfying concepts exercised therein). According to the second version of the incompatibility objection, externalism cannot accommodate the fact that rational thinkers can achieve explicatory knowledge of their thoughts and concepts through introspective reflection. I will refer to the claim that explicatory knowledge is generally available as the “explicatory access thesis”.

The explicatory access thesis forms the basis of two leading objections to externalism: the McKinsey argument and the “slow switching” argument. If content externalism is true, then a grasp of the truth conditions of one’s water thought would prima facie license an inference to the conclusion that there exists something external to oneself—appropriate watery stuff, or a community that has interacted with such stuff. Similarly, if reflection on one’s thoughts provides for a grasp of their truth conditions, a thinker who had been “switched” from Earth to Twin Earth could knowledgeably compare the truth conditions of her current “water” thought with those of a remembered “water” thought. She might thereby recognize that these thoughts differed in truth conditions: the truth conditions of the remembered thought involved water, whereas those of the current thought involve twin water. She could thus establish, without examining the environment, that her current and previous environments differed in the relevant way.

There are numerous responses to these arguments in the literature. Some of these focus on issues specific to the arguments—e.g., whether warrant “transmits” from introspective beliefs to beliefs about the external world, in the way required for licensing the anti-skeptical conclusion (Davies 1998); and whether previous concepts survive the switch to a new environment (Ludlow 1995, Tye 1998). But most responses to the

---

4 Goldberg (2003, 254) tentatively suggests that the McKinsey argument can be blocked by acknowledging that we don’t know what we are thinking, in the relevant sense—at least, that we don’t qualify as having that knowledge in the skeptical context, where the standards for knowledge may be especially high.
arguments involve acknowledging that externalism is incompatible with some kind of access to our own thought contents, and denying that we have that kind of access.

Most importantly for our purposes, many externalists—including Burge—respond to this type of objection by rejecting the explicatory access thesis. The cases Burge uses to support externalism involve subjects purported to be remarkably lacking in explicatory knowledge. His most famous subject purportedly exhibits basic ignorance about the criteria for arthritis and, hence, about the truth conditions of the thought he expresses by saying “I have arthritis in my thigh”. Burge contends that this lack of explicatory knowledge is no impediment to rational thought.

Thus, I can know that I have arthritis, and know I think I have arthritis, even though I do not have a proper criterion for what arthritis is. It is a truism that to think one’s thoughts … one must understand what one is thinking well enough to think it. But it does not follow that such understanding carries with it an ability to explicate correctly one’s thoughts or concepts via other thoughts and concepts; nor does it carry an immunity to failures of explication. So one can know what one’s thoughts are even while one understands one’s thoughts only partially, in the sense that one gives incomplete or mistaken explications of one’s thoughts or concepts. (Burge 1988, 662)

Moreover, Burge thinks that no amount of introspective reflection will correct his subject’s misunderstanding of arthritis; he can learn of his mistake only through interaction with others. So on Burge’s interpretation of the case, this subject lacks explicatory access to his thoughts.

Many externalists, including Burge, grant that externalism is incompatible with the explicatory access thesis but regard this as unproblematic because they reject that thesis. In fact, at times Burge appears to associate the explicatory access thesis with internalism. (I return to that issue in Part 2.)

1.2 Explicatory knowledge and rationality

Is the explicatory access thesis true? It’s undeniable that attempts to explicate our thoughts sometimes fall short. But the explicatory access thesis requires only that we are in principle capable of explicating any thought that we have. The most promising way to support this thesis, in the present context, is to invoke requirements for rational thought. For the point at issue between internalists and externalists is whether internal factors (always) fix a thought’s cognitive value: what it contributes to rational processes such as deliberation and inference, and how it enters into rational explanations of behavior and psychological processes. (Because this debate concerns cognitive value, I will use “content” to refer to cognitive value throughout the paper.)

One might argue that explicatory access is required for rational thought as follows. If a thought process, such as an inference, is truly rational (rather than brute), the thinker qualifies as rational for engaging in it. A thinker cannot be credited with rationality in virtue of a particular thought process unless she can understand what makes the process a rational one. What makes a thought process rational is the relations among the contents of the thoughts involved: e.g., the relations between the truth conditions of the thoughts
on which her inference is based, and the truth conditions of her conclusion. So to understand what makes a thought process rational, a thinker must have explicatory knowledge. And since such an understanding must be available, in order for the thought process to be rational, rational thought depends on the thinker’s ability to achieve explicatory knowledge.

But there are reasons to doubt that a thought’s contributing to a rational process depends on the thinker’s being in a position to explicate the thought. In fact, the debate about mental content itself casts some doubt on this claim. Externalists and internalists can presumably have, and reason with, the same thoughts. They can agree that arthritis is painful, and that water quenches thirst; and they can employ these beliefs in reasoning. What defines the debate between externalists and internalists is a disagreement about the proper explication of such thoughts. The externalist maintains that environmental factors (how community experts use “arthritis”, etc.) directly affect their contents, whereas the internalist construes their contents as fully determined by internal factors. This disagreement does not disprove the explicatory access thesis, since that thesis says only that rational thinkers must be able to explicate their thoughts. But the fact that these competing explications have withstood prolonged rational reflection makes that thesis difficult to defend.

The intuition behind the explicatory access thesis is that rational thought requires a genuine grasp of what it is that one is thinking. This idea is plausible: after all, one cannot think alligators are dangerous if one has no clue (or is radically mistaken) about what kind of worldly conditions would make this thought true or false. But we can accommodate the idea that rational thought requires genuine understanding without appealing to anything as strong as the explicatory access thesis.

A more moderate understanding of the link between rationality and understanding our own thoughts is suggested in this passage from Shoemaker.

[I]t is essential to being a rational being that one be sensitive to the contents of one’s belief-desire system in such a way as to enable its contents to be revised and updated in the light of new experience, and enable inconsistencies and incoherences in its content to be eliminated. (Shoemaker 1994, 285)

The way in which rational thinkers must be “sensitive to the contents” of their attitudes, on this description, need not involve thinking about those attitudes, or being able to explicate them. The kind of sensitivity described here may consist purely in relations between first-order states.

A simple example will illustrate this point. Rita loves to swim in the ocean, but only when it is sunny. So when she hears the weather forecaster say “tomorrow will be sunny

---

5 Shoemaker separately proposes that this sensitivity suffices for knowledge of one’s own attitudes. This is his constitutivism about self-knowledge, according to which “believing that one believes that P can be just believing that P plus having a certain level of rationality, intelligence, and so on” (ibid., 289). However, the knowledge he has in mind seems not to be the kind of higher-order knowledge at issue here. For some worries about Shoemaker’s use of this argument, see Gertler 2011, ch. 5.
and clear”, she decides to visit the beach on the following day. As she prepares to embark the next morning, she observes a dark threatening sky. The contents of her “belief-desire system” are appropriately “revised and updated in the light of [this] experience”: she abandons her plan to go to the beach, because she no longer believes that it will be sunny. Crucially, these revisions may take place entirely at the first-order level. Suppose they do. Rita abandons her previous plan upon seeing the dark clouds, without thinking about her belief that dark clouds indicate rain, or about her aversion to being at the beach when it’s raining (etc.). And for all I have said, Rita may be unable to explicate her thought dark clouds indicate rain, or to articulate how this thought contributes to rationalizing the change in plans. These revisions nonetheless seem rational in precisely the sense that Shoemaker suggests: they exhibit (or partly constitute) the requisite sensitivity to new experiences and to potential inconsistencies between attitudes.

The plausible idea that rational thought requires a grasp of what one is thinking can be accommodated without invoking the explicatory accessibility thesis. Rita’s grasp of what she’s thinking, when she thinks dark clouds indicate rain, consists in her sensitivity to its content. This sensitivity is reflected in the role that thought plays in her cognitive economy: it disposes her, upon seeing dark clouds, to abandon the belief that the day will be sunny. And a thought with a different content would play a different role.

We do sometimes reflect on, and attempt to explicate, our own attitudes. But rational revisions of our attitudes usually occur (like Rita’s) without any reflection on our attitudes or their contents. And if one can exhibit rationality without even attempting to achieve explicatory knowledge, it’s not clear why rationality should require the in-principle availability of such knowledge.

Rationality plausibly requires (or consists in) a certain kind of sensitivity to thought contents, a kind of sensitivity that constitutes thinking with comprehension. But relations among first-order states appear to suffice for the requisite kind of sensitivity. So we can make sense of the requirement that rational thinkers comprehend what they’re thinking without claiming that they are always in a position to achieve explicatory knowledge of their thoughts.

The intractability of philosophical disputes about how to explicate thought contents undermines the explicatory access thesis. And the idea motivating that thesis, that rational thought requires a meaningful grasp of what one is thinking, can be accommodated by the kind of first-order sensitivity that Shoemaker plausibly describes as characteristic of rationality. So the externalist is on firm footing in denying that thinkers generally possess (or can acquire) explicatory knowledge. Externalism is not threatened by its alleged incompatibility with the explicatory access thesis.

The issue of whether externalism is compatible with thinkers’ access to the proper explication of their thoughts is a red herring. The truth of externalism instead depends on whether external factors directly affect rational relations among first-order thoughts and attitudes. Part 2 takes up this question.
PART 2

2.1 Differential dubitability

The question dividing externalists and internalists is whether external factors directly affect a thought’s cognitive value: that feature by virtue of which it stands in rational relations, such as inconsistency or entailment, to other thoughts. The lesson of Part 1 is that, to evaluate this question, we should direct our attention to the primary locus of rationality, the relations between first-order mental states.

I will approach this question by using a famous Fregean thesis about cognitive value. According to Frege, the thought that \( p \) and the thought that \( q \) differ in cognitive value if a rational thinker can believe that \( p \) while doubting that \( q \).\(^6\) In other words, differential dubitability implies a difference in cognitive value. It is significant that this thesis does not entail that a rational thinker will be able to explicate her thoughts, or that she can ascertain, through reflection, \( that \) two of her thoughts share a cognitive value. This Fregean thesis directly involves only the first-order sensitivity described in the previous section. It is precisely because rational thinkers are sensitive to the cognitive values of their thoughts that a rational thinker could not differentially doubt two thoughts that shared a cognitive value.

Frege drew on this thesis to explain how a rational thinker could believe that \( \text{Hesperus is Hesperus} \) while doubting that \( \text{Hesperus is Phosphorus} \), despite the fact that \( \text{Hesperus} \) and \( \text{Phosphorus} \) are co-referential. That these thoughts differ in cognitive value demonstrates that cognitive value can vary independently of actual reference (and of the truth conditions associated with actual reference).

The Fregean thesis that differentially dubitable thoughts differ in cognitive value may appear to conflict with externalism. Consider an example given by Brian Loar. A child named Paul says that “chats have tails”; this expresses a belief he formed on the basis of what his French nanny tells him about the cats around him. On the rare occasions that Paul sees his parents, cats are never present; yet they tell him about animals called “cats” and, in particular, that “cats have tails”. Because Paul doesn’t realize that “chat” and “cat” are co-referential, the beliefs he expresses with “chats have tails” and “cats have tails” are differentially dubitable (for him). So, according to the Fregean thesis, they differ in cognitive value. But the external factors that contribute to the cognitive value of \( \text{chats have tails} \)—such as the social practice of using “chat” to refer to felines—are precisely those that contribute to the cognitive value of \( \text{cats have tails} \). So one might conclude that externalism will imply that these \( \text{share} \) a cognitive value.

But externalism is not directly threatened by this Fregean thesis. The externalist does not claim that it is \textit{only} external factors that individuate contents; externalism is compatible with the claim that a difference in \textit{internal} factors can suffice for a difference in contents. So the externalist can attribute the differential dubitability of the thoughts

---

\(^6\) In my use, “\( p \)” and “\( q \)” represent propositional contents of thought, contents which can be entertained at different times and by different thinkers. I remain neutral as to whether such contents are best understood as abstract propositions, modes of presentation of propositions, sentences in a language of thought, etc.
Paul expresses with “chats have tails” and “cats have tails” to a difference in cognitive value between these, a difference that is anchored in factors internal to Paul. And it is plausible that these thoughts do differ internally, given the disparate etiologies of Paul’s “chat” and “cat” concepts. More generally, since externalism does not claim that cognitive value is individuated by external factors alone, it is not directly threatened by cases in which cognitive value varies independently of external factors. So the Fregean thesis mentioned above, that differential dubitability implies a difference in cognitive value, poses no direct threat to externalism.⁷

This point does not resolve the main worry about externalism, that the cognitive values of our thoughts seem independent of external factors. It merely reveals the proper expression of that worry. The problem for externalism is not that differential dubitability is sufficient for a difference in content, but rather that it is necessary for a difference in content.⁸ Wikforss (2006) puts this succinctly:

> The content externalist holds that thought content is individuated by the individual’s environment in such a way that two individuals may be internally identical and yet have different thoughts. This thesis implies that there may be a difference in content that cannot be discerned from the first person point of view, but it does not imply that there may be sameness of content that is not discernible from a first person point of view. (Wikforss ibid., xx, notes elided)

What I have described as the main worry about externalism is best pursued by focusing on the implication that, in Wikforss’s terms, “there may be a difference in content that cannot be discerned from the first person point of view”. As I will put this, it is that there may be a difference in content to which rational thinkers cannot be sensitive (in the sense discussed above). Developing an objection to externalism along these lines requires supplementing our Fregean thesis about cognitive value.

---

⁷ Boghossian (1994) uses the cat-chat argument to show that externalism is incompatible with a thesis he calls “epistemic transparency”:

> [F]or any two expressions in a person’s repertoire, it must be possible for that person to come to know whether or not they mean the same … purely introspectively, without the benefit of further empirical investigation”. (Boghossian 1994, 33)

Note that this closely resembles the explicatory access thesis, which Burge and others reject. But Boghossian regards epistemic transparency as required for rationality. He says “It is only by attributing epistemically transparent thoughts, to a subject, that we can assess the subject’s rationality and construct rational explanations of her behavior.” (ibid. 39) While I cannot do justice to his argument here, I think the case of Rita casts doubt on the idea that thinkers assessable for rationality must have reflective access to their thoughts, as required by the epistemic transparency thesis.

⁸ The terms “sufficient” and “necessary” are used, as is standard, to express purely logical relationships. So the thesis that differential dubitability suffices for a difference in cognitive value is neutral as to whether differential dubitability reflects or constitutes (etc.) this difference.
2.2 The differential dubitability test

Cognitive value is that aspect of a thought that bears on rational assessments of the thinker or, perhaps equivalently, of her psychological profile (relations among her thoughts, attitudes, and experiences). A way of individuating thoughts by their cognitive values must therefore meet two conditions. It must be fine-grained enough to capture differences between cognitive values to which a rational thinker could be sensitive. This requirement is reflected in Frege’s thesis that differential dubitability—which reflects first-order sensitivity—suffices for a difference in cognitive value. But it must also be coarse-grained enough to ignore distinctions beyond those to which a rational thinker could be sensitive. This second condition implies that, for a restricted range of thoughts, differential dubitability is also necessary for a difference in cognitive value. It is this latter requirement that most directly challenges externalism.

The idea that differential dubitability is necessary for a difference in cognitive value faces immediate obstacles. One problem is that there may be some thoughts that a rational thinker cannot doubt, and others that a rational thinker cannot believe. Unless we are willing to say that there is a single cognitive value shared by all indubitable thoughts (such as, perhaps, \(2+3=5\) and \(I\ exist\)), and another shared by all unbelievable thoughts (e.g., \(2+3=7\) and \(I\ do\ not\ exist\)), the requirement of differential dubitability must be restricted in its scope.

I propose that the differential dubitability of the thought that \(p\) and the thought that \(q\) is necessary for a difference in cognitive value so long as the pair \((p, q)\) meets three restrictions.

(i) \(p\) and \(q\) are each rationally dubitable and rationally believable; and neither \(p\) nor \(q\) are conjunctions that include an indubitable or unbelievable conjunct.  

(ii) Neither \(p\) nor \(q\) concerns one’s occurrent thoughts.

(iii) ‘\(p\ iff q\)’ is not knowable a priori.

The latter clause excludes contents like no bachelors are married and snow is white and \(2+3=5\) and snow is white. These are each believable and dubitable, but the pair may not be differentially dubitable although they plausibly differ in cognitive value.

The problem with thoughts about thoughts stems from the possibility that thoughts are self-intimating, a possibility on which I remain neutral. If thoughts are self-intimating, then it will be impossible to believe I am having exactly one occurrent thought right now while simultaneously doubting I am having exactly one occurrent thought right now and snow is white. For if thoughts are self-intimating, it is not possible to have both occurrent thoughts while believing that one has only a single occurrent thought. (I am grateful to Alan Sidelle for bringing this point to my attention.)

Condition (iii) may introduce a worry of triviality: it seems trivial that all distinct contents meeting these conditions will be differentially dubitable. But the cases to which I apply this test are precisely cases that (according to the externalist) meet conditions (i)-(iii), but are not differentially dubitable.

---

9 The latter clause excludes contents like no bachelors are married and snow is white and \(2+3=5\) and snow is white. These are each believable and dubitable, but the pair may not be differentially dubitable although they plausibly differ in cognitive value.

10 The problem with thoughts about thoughts stems from the possibility that thoughts are self-intimating, a possibility on which I remain neutral. If thoughts are self-intimating, then it will be impossible to believe I am having exactly one occurrent thought right now while simultaneously doubting I am having exactly one occurrent thought right now and snow is white. For if thoughts are self-intimating, it is not possible to have both occurrent thoughts while believing that one has only a single occurrent thought. (I am grateful to Alan Sidelle for bringing this point to my attention.)

11 Consider the case where \(p = \text{‘this figure is a right triangle’}\), and \(q = \text{‘the square of the length of this figure’s hypotenuse is equal to the sum of the squares of the length of its other sides’}\). Since it is a priori that \(p\ iff q\), a rational thinker may not be able to believe that \(p\) while doubting that \(q\) (or vice versa). Condition (iii) may introduce a worry of triviality: it seems trivial that all distinct contents meeting these conditions will be differentially dubitable. But the cases to which I apply this test are precisely cases that (according to the externalist) meet conditions (i)-(iii), but are not differentially dubitable.
I remain neutral as to whether each of these restrictions is needed. My aim is not to defend a general Fregeanism, but only to develop and apply a broadly Fregean strategy for differentiating cognitive value within the realm in which it is least problematic: viz., among pairs of contents that meet the restrictions just outlined. My proposal is as follows.

**The Differential Dubitability Test**

For any pair of thought contents \((p, q)\) that meet conditions (i)-(iii):

\(p\) and \(q\) differ in cognitive value IFF a rational thinker (who can entertain \(p\) and \(q\)) can believe that \(p\) while doubting that \(q\) OR believe that \(q\) while doubting that \(p\).

The differential dubitability test extends the Fregean thesis discussed earlier, and shares its motivating idea: that differences in cognitive value are precisely those to which rational thinkers can be sensitive.

The differential dubitability test is at odds with Burge’s externalism. The subject in Burge’s “arthritis” case, who is stipulated to be rational, has a belief he expresses by saying “I have arthritis in my thigh”. Burge says that, while this belief is false, the subject would have a *true* belief in a counterfactual situation in which he retains all of the same (non-intentionally specified) internal properties and history, but in which “physicians, lexicographers, and informed laymen apply ‘arthritis’ not only to arthritis but to various other rheumatoid ailments.” (Burge 1979, 105) Given that the subject would have a true belief in that situation, it seems that the subject in the actual situation cannot differentially doubt *I have arthritis in my thigh* and *I have a rheumatoid ailment (of the relevant, more inclusive kind) in my thigh*. (To avoid worries about whether the subject could entertain the latter thought, we can suppose that he has been switched, without his knowledge, from an environment that fits the counterfactual situation Burge describes.) And as Burge construes these thoughts, they meet conditions (i)-(iii). So according to the differential dubitability test, these share a cognitive value. But the principal thesis of Burge’s externalism is precisely that these thoughts differ in cognitive value. So Burge’s externalism conflicts with the differential dubitability test.

### 2.3 Burge’s proposed modification to the differential dubitability test

Burge recognizes that his externalism conflicts with the differential dubitability test. But he does not reject Fregeanism entirely, perhaps because he regards cognitive value as closely tied to differential dubitability. In an effort to reconcile externalism with what is intuitive about Fregeanism, he proposes a modification to the differential dubitability test. “Frege’s test for differential dubitability, *when accompanied with requirements that doubt be supportable by publicly recognized methods*, is a defeasible but profoundly valuable tool in individuating cognitive values.” (Burge 1986, 717; my emphasis).

Burge’s proposal, that doubt must be publicly supportable, seems to target the claim that differential dubitability by an individual suffices (on its own) for a difference in cognitive value. But as we saw above, the conflict with externalism derives not from this claim but, rather, from the claim that differential dubitability is *necessary* for a difference in cognitive value. To avoid this conflict, the externalist must allow that two thoughts may
differ in cognitive value even if a rational thinker is not able to differentially doubt them. Burge’s proposed modification supports this position if it is also applied, as seems reasonable, to the claim that differential dubitability is necessary for a difference in cognitive value. When thus modified, the differential dubitability test says that \( p \) and \( q \) differ in cognitive value iff “publicly recognized methods” would support differential dubitability.

The modified differential dubitability test is compatible with externalism. For example, the differential dubitability of *I have arthritis in my thigh* and *I have a rheumatoid ailment in my thigh* is clearly “supportable by publicly recognized methods” of discriminating arthritis from other rheumatoid ailments. So these differ in cognitive value, according to the modified differential dubitability test.

Burge’s modification introduces an external factor into the test for sameness and difference of cognitive value. If this modification is needed to ensure the (relative) accuracy of the differential dubitability test, as Burge suggests, then external factors can directly affect cognitive value. Given the close tie between cognitive value and differential dubitability, the debate between internalism and externalism might well be decided by the question whether the original differential dubitability test has flaws—e.g., it sometimes delivers the wrong result—and that these flaws are remedied by Burge’s proposed modification.

What reason might there be to think that the original differential dubitability test is flawed?

Externalist suspicions about the original differential dubitability test appear to stem from the perception that it overestimates thinkers’ access to their own thought contents. After all, the most obvious examples of the original differential dubitability test delivering the right results are cases in which the contents of two thoughts, or the relations between them, are unusually accessible. For instance, a rational thinker will be able to differentially doubt *some Spaniards play tennis* and *some Spaniards are 8 feet tall*, the original differential dubitability test delivers the right result in this case only because those with the requisite concepts generally have enough explicatory knowledge to recognize that “all and only tennis players are 8 feet tall” does not express a criterial truth.

But as I argued above, we are often ignorant of (and/or mistaken about) criterial truths that contribute to the cognitive values of our thoughts. Burge repeatedly emphasizes this point. His view seems to be that this lack of explicatory knowledge prevents differential dubitability from tracking differences in cognitive values. To adapt one of his examples (Burge 1986): it appears that a rational person may be unable to differentially doubt *I’m seated on a chair* and *I’m seated on a piece of furniture with legs, intended for sitting*. But these thoughts do differ in truth conditions: if I were seated on a ski-lift chair, the former would be true while the latter was false. The implication is that what one can differentially doubt is influenced by criterial beliefs (e.g. the mistaken criterial beliefs that chairs must have legs), whereas cognitive value has to do with the actual criteria for *chair*. Burge thus seems to think that the influence of mistaken criterial beliefs is enough to render the original differential dubitability test unreliable.
The original differential dubitability test conflicts with externalism. Burge objects to the original differential dubitability test, seemingly on the grounds that mistaken criterial beliefs can prevent differential dubitability from tracking difference of cognitive value. His proposed modification is intended to avoid the conflict while preserving a measure of Fregeanism.

2.4 Defending the original differential dubitability test

The original differential dubitability test does not, I think, deliver the wrong results in cases of mistaken criterial beliefs. In fact, I will show that it is accurate in an extreme case, involving a subject Burge describes who is not only ignorant of key criterial truths but also has radically misguided criterial beliefs. The externalist construal of such cases, as counter-examples to the original differential dubitability test, rests on a confusion between the differential dubitability phenomena and explicatory knowledge of thoughts.

The case I have in mind is the case of Burge’s subject, A, who thinks that sofas are religious objects. A believes that what makes something a sofa is that it is a religious artifact of a certain sort. A recognizes that others regard sofas as intended for sitting as a criterial truth, but he thinks that they are mistaken. Clearly, A has inaccurate criterial beliefs about sofa. And in failing to believe that sofas are intended for sitting, he is ignorant of a key criterial truth about sofa.

Consider A’s position relative to these two thoughts.

(r) The green item in the corner is a certain sort of religious artifact.

(s) The green item in the corner is a sofa.

Recall that the objection to the original differential dubitability test is that mistaken criterial beliefs can affect what is differentially dubitable for an individual. According to that objection, A’s religious theory of sofas will prevent him from believing (s) while doubting (r). Further supposing that A takes sofas to be the only religious artifacts of the relevant sort, A will be similarly unable to believe that (r) while doubting (s).

The challenge to the original differential dubitability test is clear. Thoughts (r) and (s) seem plainly to differ in cognitive value, yet they are not differentially dubitable for A (who is stipulated to be rational). This appears to be a counter-example to the component of the original differential dubitability test that threatens externalism, namely, the claim that rational thinkers will be able to differentially doubt any two thoughts that differ in cognitive value. Moreover, Burge’s proposed modification to that test remedies this flaw. Since the differential dubitability of (r) and (s) is clearly “supportable by publicly recognized methods”, these differ in cognitive value, according to the modified test.

12 Burge uses the case of A for a different purpose; I discuss his use of it below.

13 Burge describes A’s views as “nonstandard theory”. But they are clearly (also) criterial beliefs, as they are beliefs about what makes something a sofa. “A doubts the truisms and hypothesizes that sofas function not as furnishings to be sat on, but as works of art or religious artifacts. … A admits that some sofas have been sat upon, but thinks that most sofas would collapse under any considerable weight and denies that sitting is what sofas are pre-eminently for.” (Burge 1986, 707)
However, this challenge to the original differential dubitability test rests on an unsustainable interpretation of the case of $A$. Given $A$'s unorthodox views about the things he calls “sofas”, it is not obvious that he has the concept sofa, as required for entertaining $r$. To establish that $A$ possesses the concept sofa, Burge stipulates that $A$ is willing to submit his religious theory to empirical testing, and to relinquish that theory if it is disproven according to communal norms of evidence. “We may imagine that if we were to carry out [A’s] proposed experiments [meant to establish that sofas are religious artifacts], $A$ would come to admit that his theory is mistaken.” (ibid., 263) So although $A$ does not defer to others’ criterial beliefs about sofa, he does defer to communal epistemic norms about how to determine what makes something a sofa.

The stipulation that $A$ is willing to relinquish his religious theory, under those conditions, justifies crediting him with the concept sofa. But it also defuses the challenge to the original differential dubitability test. For it ensures that $A$ can differentially doubt ($r$) and ($s$) after all. $A$ can believe that ($s$) while doubting that ($r$) precisely by imagining that his proposed experiments have disproven his religious theory of sofas. (Alternatively, we could bring him to believe this by actually conducting the experiments or duping him into believing that we have.)

This defense of the original differential dubitability test cannot be blocked simply by revoking the stipulation that $A$ is willing to abandon his religious theory under certain conditions. Suppose that $A$’s adherence to the belief that sofas are religious artifacts were unshakeable. If $A$ were disposed to maintain his religious theory come what may, then $A$’s “sofa” concept would not be sofa but, rather an idiosyncratic concept that he mistakenly associates with “sofa”. (The actual extension of this idiosyncratic concept would presumably be empty.) Lacking the concept sofa, $A$ could not entertain ($s$) at all. So $A$’s willingness to abandon his religious theory is both necessary for his possessing the concept sofa and, assuming that he has the other concepts required to entertain ($r$) and ($s$), sufficient for him to differentially doubt ($r$) and ($s$).

I have focused on $A$’s radically mistaken criterial beliefs. Burge uses this case for a different purpose: to show that the argument for externalism does not rest on factors present in his earlier arguments, viz., “incomplete linguistic understanding and ignorance of expert knowledge” (ibid., 708). But the argument involving $A$ shares a central characteristic of those earlier arguments. All of these arguments involve a thinker who fails to fully understand a given concept, and defers to an external (usually social) factor about that concept. $A$ is unusual in that he does not defer to the beliefs of his community, or experts therein, about what makes something a sofa. He defers to the results of his proposed experiments. But it is this deference which constitutes his willingness to relinquish his religious theory, and thereby allows him to be credited with the concept sofa.

Burge’s externalist arguments involve rational thinkers who are purported to exercise concepts in thought, while harboring mistaken criterial beliefs about those
concepts (or exhibiting other sorts of explicatory errors or ignorance). These thinkers cannot correct their errors or cure their ignorance through introspective reflection: they can be corrected only by external factors (communal usage, local experts, the outcome of empirical testing, etc.). These cases thus appear to be counter-examples to the explicatory access thesis.

But these cases do not directly threaten either the original differential dubitability test or internalism. They do not threaten the original differential dubitability test because explicatory beliefs, including criterial beliefs, are beliefs about one’s own thoughts and concepts, whereas differential dubitability is a matter of first-order sensitivity: how one reasons with a thought, and how employs the concepts exercised therein. The original differential dubitability test is neutral on the question whether rational thinkers can achieve reflective, explicatory knowledge about their thoughts; it is thus neutral about whether they can recognize that two of their thoughts differ in (or share a) cognitive value.

Burge’s cases do not threaten internalism because internalism is not accessibilism: it is neutral about whether thinkers have explicatory access to their thoughts. Internalism holds that the cognitive values of thoughts are fixed by factors internal to the thinker. In other words, internal factors fix the rational role of a thought that $p$—the conditions under which one is disposed to believe that $p$, this thought’s rational relations to other thoughts, the kind of inferences into which it enters, etc. This role is (at least primarily) a matter of relations between first-order states. Internalism is neutral as to whether any relations to higher-order states are implicated in rationality. Internalism is true so long as the cognitive value of thoughts supervenes on factors within the thinker, regardless of whether the thinker has reflective access to the proper explication of those thoughts. Cognitive value is largely a matter of dispositions, including dispositions that constitute deference to external factors. (The internal factors on which cognitive values supervene may include categorical grounds of these dispositions.)

2.5 Deference and indexicality

Burge’s externalist arguments highlight a previously neglected aspect of our dispositions to apply certain concepts: deference. (It shares this feature with the anti-descriptivist arguments of Putnam (1975) and Kripke (1980).) The objects of deference can be social or physical; what is important to Burge’s externalist arguments is that we routinely defer to factors external to ourselves.

Internalism can accommodate the phenomenon of deference: the internalist will construe deferential concepts as indexed to the social or physical factors to which the

---

14 This is not to say that external factors are irrelevant in cases where the individual knows the salient criterial truths. Since external factors contribute to fixing criterial truths, according to externalism, the truth of the subject’s criterial beliefs depend, in part, on external factors.

15 The idea that the cognitive value of a thought is fixed by the individual’s dispositions, including dispositions that constitute deference to communal norms, is a familiar one. This idea forms the basis of semantic two-dimensionalism, which is internalist about the cognitive value of thoughts (what Chalmers calls their primary intension) while allowing that other kinds of content (including Chalmers’ secondary intensions) depend on external factors.
thinker defers. Burge criticizes this strategy at numerous places; here are two representative passages.

‘Sofa’ in our attributions to A is not indexical. … One could perhaps modify the italicized phrase to thing of a kind relevant to understanding what I (or we) usually refer to when I have this sort of experience (or, alternatively, when I use ‘sofa’). But now it is clear that no one but a philosopher would think of sofas in that complex, meta-level way. A complex analysis or theory of reference should not be conflated with the way A thinks of sofas. (Burge 1986, note 12)

Sometimes arthritis is called a ‘deferential concept’. This phrase seems to me very misleading. Nearly any concept can be employed in such a way that the employer depends on others for the range of the concept’s application, and even for instruction on explicational principles and other norms governing the concept. Our reliance on others places us under standards and norms that we may not have fully mastered. Moreover, we cannot in general tell by simple reflection whether and how we depend on others. The dependence commonly is buried in the history of one’s usage and in dispositions not all of which are open to reflective recognition. The main issue has to do with what objective reality we are connected to and what standards for full understanding apply to those aspects of our usage that rely on such connection. (Burge 2006, PS to I&T, 173)

These passages are illuminating. Burge appears to regard the claim that sofa or arthritis are indexical concepts, as implying that the subject can discover this fact through introspective reflection. But it does not have this consequence unless the explicatory access thesis is true. The internalist can remain neutral on that thesis. She can thereby maintain that my arthritis concept is indexed to social factors to which I defer, without committing herself to the claim that I can recognize that it is deferential. Thought contents may supervene on internal factors even if thinkers cannot always explicate them, through introspective reflection; and even if a rational thinker was unable to recognize the correct “complex, meta-level” analysis of her arthritis concept. To assume that the correct analysis of a concept does require the availability of such knowledge is to assume the explicatory access thesis. But internalism is independent of accessibilism.

Why might Burge (or anyone) think that these are connected? The answer might lie in an implicit conception of what is “internal to the thinker”. As I have recently argued (Gertler 2012), there is no univocal reading of this phrase, and hence no single boundary of the thinker, operative in the current debate over content. Burge’s criticisms of the idea that concepts like arthritis are narrow, indexical concepts, would be justified if “thoughts and concepts supervene on factors internal to the thinker” implied “thoughts and concepts are explicable by the thinker”. So although Burge’s explicit characterizations of what makes a factor internal or external are not epistemic, his criticisms of the internalist construal of deferential concepts may rest on an implicit conception of what is “internal
to the thinker” as what she can explicate. That is, these criticisms seem implicitly to construe internalism as accessibilism.\footnote{Farkas (2003) supports a related construal of internalism; I express some worries in my 2012 paper.}

While Burge rightly emphasizes the difference between having a particular thought and being able to explicate that thought, he neglects some of the possibilities this difference opens up. For his discussion implies that we face a stark choice: either allow that cognitive value depends on factors external to the thinker, or claim that individual subjects always have access, through reflection, to the proper explication of their thoughts. But there is a third option. The cognitive value of a thought is fixed, in part, by the subject’s implicit dispositions—to make certain inferences, to appropriately revise other attitudes in light of the thought (etc.). The cognitive value of a concept—what it contributes to the rational role of thoughts—is fixed, in part, by the thinker’s implicit dispositions to apply the concept to certain things and not others, and to adjust her use according to external factors to which she defers. Rational thinkers need not know that they are so disposed; that a concept is indexed to external factors to which the thinker defers does not guarantee that the thinker can recognize this fact.

A crucial component in Burge’s arguments for externalism (and, relatedly, against the original differential dubitability test) is the idea that thinking doesn’t require full understanding, or avoiding misunderstanding. These arguments threaten the expiatory access thesis. That thesis is further threatened by the arguments in section 1.2 above. And the intuition supporting it can be captured in the notion I extracted from Shoemaker: that the comprehension of content required for rational thought is a matter of first-order sensitivity. Because internalism does not entail accessibilism, the internalist need be no more committed to the expiatory access thesis than the externalist is. Internalism requires only that the rational relations between first-order thoughts are grounded in factors internal to the thinker. Burge’s arguments show that individuals routinely lack expiatory knowledge, and defer to external factors about the proper application of their concepts. But while these conclusions cast doubt on the expiatory access thesis and, hence, on accessibilism, they pose no direct challenge to internalism.