SAFETY, CONTENT, APRIORITY, SELF-KNOWLEDGE*

This essay motivates a revised version of the epistemic condition of *safety* and then employs the revision to (i) challenge traditional conceptions of apriority, (ii) refute "strong privileged access," and (iii) resolve a well-known puzzle about externalism and self-knowledge.

I. VARIETIES OF EPISTEMIC RISK

I think I hear a lark outside. Suppose I am right, but my belief could easily have gone amiss. In that case I do not know that I hear a lark. For short: if I know that *p*, I must "safely" believe that *p*. But how might a belief go amiss? Let me count the ways.

First, I might have falsely believed that I hear a lark. If this possibility threatens—for example, there are many lark-imitating imposters nearby—then I do not know that I hear a lark. As safety is usually formulated, it only rules out this type of threat:

STANDARD SAFETY: *S* could not easily have falsely believed that *p*.[1]

But safety should do more. A second way I could have messed up is by falsely believing some other, closely related proposition.[2] Suppose I have a true demonstrative thought, one that I might express by saying, 'That is a lark'. If the imposters are nearby, I am still in danger

* Many of the ideas in this paper, especially in section III, grew out of conversations with John Hawthorne: I owe a great deal to him. Warm thanks also to Ryan Wasserman, Brian Weatherson, and Takashi Yagisawa for discussion and comments on earlier versions of this paper.

[1] For a sampling of those who endorse some version of safety as a necessary condition on knowledge, see Stewart Cohen, "Contextualism and Skepticism," *Philosophical Issues*, X (2000): 94–107; Keith DeRose, "Solving the Skeptical Problem," *Philosophical Review*, CIV (1995): 1–52; R.M. Sainsbury, "Easy Possibilities," *Philosophy and Phenomenological Research*, LVII (1997): 907–19; Ernest Sosa, "Relevant Alternatives, Contextualism Included," *Philosophical Studies*, CXIX (2004): 35–65; and Timothy Williamson, *Knowledge and Its Limits* (New York: Oxford, 2000).

[2] Sainsbury, *op. cit.*, and Brian Weatherson, "Luminous Margins," *The Australasian Journal of Philosophy*, LXXXII (2004): 373–83, make a persuasive case that safety should not merely rule out close worlds where the very same proposition is believed falsely. And such a version of safety is required by the epistemic theory of vagueness in Williamson, *Vagueness* (New York: Routledge, 1994). For other excellent discussions of safety and content-switching involving demonstratives, see also Jessica Brown, "Reliabilism, Knowledge, and Mental Content," *Proceedings of the Aristotelian Society*, C (2000): 115–35, and James Pryor, "Comments on Sosa's 'Relevant Alternatives, Contextualism Included'," *Philosophical Studies*, CXIX (2004): 67–72.

0022-362X/07/0408/403–23

of messing up. But the problem is not that I could easily have falsely believed that very proposition. (Arguably it is a necessary truth.[3]) Had an imposter been singing, I would have believed something else—a proposition with different truth conditions—though I would have expressed my belief the same way. Since standard safety only rules out close possibilities in which I falsely believe that *p*, the demonstrative belief counts as safe even though the general belief that I hear some lark or other does not. But both beliefs intuitively fail to count as knowledge. After all, it is not an epistemically risky undertaking to infer the general truth from the demonstrative one.

Yet a third way I might have messed up is by having a thought with gappy content, paradoxical content, or no content at all.[4] Suppose that (unwittingly) I often hallucinate lark calls. In between some hallucinations, I hear a real lark and form the demonstrative thought I express by 'That bird is a lark'. According to a leading semantic picture, if I had been hallucinating instead, my thought episode would have had no content, because only a singular proposition would do. In that case, I would have been messing up, even though I would not have believed the proposition I actually do, and the thought I give voice to by 'That bird is a lark' would not have been false. (The same point can be illustrated by assuming semantic externalism for predicates.[5])

An alternative view has it that, had I been hallucinating, my thought would have had *gappy* content. If gappy propositions have no truth value,[6] then intuitively one messes up by believing them: we

[3] I assume that the content of a demonstrative thought is a proposition involving some object *a* whose truth value with respect to any possible world depends solely on how things stand with *a* at that world. This leaves it open whether, in a successful case, the semantic value of a demonstrative is simply its referent. One can achieve the relevant kind of object dependence with a view along the lines of John McDowell, "De Re Senses," *Philosophical Quarterly*, XXXVI (1984): 281–94, or even with a quantificational approach to demonstratives: see Jeffrey C. King, *Complex Demonstratives* (Cambridge: MIT, 2001), and Hawthorne and Manley, *Something in Mind* (New York: Oxford, forthcoming), chapter 6.

[4] Brown, *op. cit.*, notes that "local reliability" (in no nearby situation is *p* false and *S*'s method produces the belief that *p*) allows demonstrative thoughts to be known even in cases where the subject is prone to empty thoughts.

[5] Suppose that experts settle which precise property we express by 'lark'. And suppose that they nearly adopted a more exclusive membership condition, one according to which the bird I hear would not have fallen under the extension of 'lark'. Then the belief I would have expressed with 'That is a lark' would have been a false belief involving a closely related property that the demonstrated bird does not have. (Here nothing special hinges on the fact that 'lark' is a natural kind term: semantic deference is a source of externalism even for functional kind terms: for example, 'diode' and 'Wankel rotary engine'.)

[6] See Nathan Salmon, "Nonexistence," *Noûs*, XXXII (1998): 277–319.

aim at truth with our beliefs. If instead we consider all atomic gappy propositions to be false and their negations true,[7] we should hold that a subject messes up by believing even *true* gappy propositions—and safety will need revision to account for that fact. For I am failing equally, from an epistemic point of view, whether I point at a dark empty corner and say 'He is fat', or I point at a dark empty corner and say 'He is not fat'. Supposing (contrary to intuition) that the second sentence counts as true, we should still consider the associated demonstrative thought to be an epistemic failure of the sort whose nearness and similarity undermines knowledge. In short, the danger of emptiness—no less than the danger of falsehood—dispels knowledge.

This can be illustrated without appealing to object-dependence or social externalism; consider instead the phenomenon of paradox. Suppose Mary believes that what Susan just wrote in her diary is false. Aside from this belief, all of Mary's beliefs about Susan are true. It turns out that Susan wrote: 'Everything Mary believes about me is true'. As things stand, Mary's belief is paradoxical: we cannot evaluate it for truth or falsehood without contradiction. But had Susan written 'I had eggs for breakfast' instead, Mary's belief would have been truth evaluable. This is the sort of case about which Saul Kripke wrote: "Our statements involving the notion of truth [are] risky: they risk being paradoxical if the empirical facts are extremely (and unexpectedly) unfavorable. There can be no syntactic or semantic 'sieve' that will winnow out the 'bad' cases while preserving the 'good' ones."[8] Whether we say that Mary's thought in the bad case fails to have propositional content, or that its content simply cannot be evaluated for truth,[9] having a paradoxical thought is a way of messing up.

Clearly, thoughts can be in imminent danger of paradox while not being in any danger of falsehood: suppose Susan actually wrote something false, had no inclination to write anything true, but nearly wrote something paradoxical. Then Mary's belief is true, and she is not in danger of having a false belief, but she is in danger of epistemic failure. There are even beliefs that can be paradoxical but cannot be untrue: for example, the belief that someone is thinking something

[7] This is the view of David Braun, "Empty Names," *Noûs*, XXVII (1993): 449–69. Note that on this approach the gappy proposition expressed by 'Vulcan is not Vulcan' is true; moreover it is the same proposition as that expressed by 'Vulcan is not Santa Claus'. See Ben Caplan, "Empty Names," in *The Encyclopedia of Language and Linguistics* (New York: Elsevier, 2006, 2nd ed.), pp. 132–36; and also Anthony Everett, "Empty Names and 'Gappy' Propositions," *Philosophical Studies*, CXVI (2003): 1–36.

[8] Kripke, "Outline of a Theory of Truth," this JOURNAL, LXXII, 19 (November 6, 1975): 690–716.

[9] As recommended by, for example, Scott Soames, *Understanding Truth* (New York: Oxford, 1999).

untrue right now.[10] (If someone else is thinking something untrue, then it is true. If not, it is paradoxical.) It should not follow that this belief is always safe. Suppose someone tokens it in a very sparsely populated world in which people spend much of their time in deep sleep. If a belief could easily have been a Liar without being discarded by the subject, it is not knowledge.

II. SAFETY MEASURES

Since standard safety is normally advanced as a necessary condition for knowledge, these cases are not strictly counterexamples. The point is that it arbitrarily focuses on the threat of error about *p*, when threats of error involving other propositions, as well as threats of emptiness and paradox, can undermine one's knowledge that *p*.

Let us call any thought a *failed thought* if it has no content, or has as its content a false, gappy, or paradoxical proposition. It is tempting to reformulate the safety condition by simply requiring that the very thought token whereby *S* believes that *p* could not easily have been a failed thought.[11] But this hangs too much on the individuation conditions of thoughts; for instance, we would require the contentious assumption that contentful thought tokens do not have their contents essentially. Without preempting these metaphysical issues, how can we clarify the idea of the near danger of "messing up?" We need a counterpart relation on thought, likely broader than identity, such that *S* knows that *p* only if:

REVISED SAFETY: *S* could not easily have had a failed counterpart thought.

The idea is that one is safer the more distant are the closest worlds in which one has a failed counterpart thought. Revised safety sets a threshold of closeness such that no belief can be knowledge if there are false or empty counterpart thoughts in worlds at least that close. For those who are drawn to contextualism in epistemology, it is natural to hold that the threshold degree is context sensitive.[12]

Note that we need a threshold for closeness of worlds *as well as* a counterpart relation for thoughts. Very different untrue thoughts at

[10] The example is from John Hawthorne, "Deeply Contingent A Priori Knowledge," *Philosophy and Phenomenological Research*, LXV (2002): 247–69.

[11] See the notion of "belief-safety" in Weatherson, *op. cit.*

[12] For gradable versions of standard safety with a contextually specified threshold, see: DeRose, "Sosa, Safety, Sensitivity, and Skeptical Hypotheses," in John Greco, ed., *Sosa and His Critics* (Malden, MA: Blackwell, 2004), pp. 22–41; Sosa, "Relevant Alternatives, Contextualism Included"; and Mark Heller, "The Proper Role for Contextualism in an Anti-Luck Epistemology," *Philosophical Perspectives*, XIII (1999): 115–29. If closeness only involves a partial ordering of worlds, the comparative will sometimes lapse.

very close worlds do not count against knowledge; and neither do very similar untrue thoughts in very distant worlds. We could gerrymander a similarity* relation for thoughts that builds in closeness of worlds, so that two duplicate thoughts in different worlds could differ in their similarity* to a third thought. But the resulting concision is artificial, concealing as it does two quite different elements contributing to the notion of safety: one having to do with similarity between thoughts; the other with overall similarity of worlds.[13]

I will not attempt a precise analysis of the needed counterpart relation for thoughts; indeed, the vagueness of 'knows that *p*' may not allow precision in this endeavor. But there are some clear structural constraints on the relation. First, if empty thoughts can be counterparts of contentful thoughts, we cannot characterize the relation in terms of similar content. Second, there is pressure to deny that two beliefs are counterparts if they are based on very different evidence. Consider the following problem for standard safety discussed by Ernest Sosa.[14] Suppose I see a large fish surface. I recognize the distinctive dorsal fin of a sailfish, and I infer that a fish has surfaced nearby. But a whale almost surfaced instead, and had the whale surfaced I would have thought falsely that a fish had surfaced. In this case, while my actual belief fails the standard safety test, it seems to be a piece of knowledge—as long as I never mess up when confronted with sailfish-evidence.

To handle this problem, Sosa suggests a version of safety along these lines:

> BASIS: *S* could not easily have believed *p* on the same basis without *p* being true.[15]

But this condition, like standard safety, does not rule out the threats at issue in section I. Since the proposition I express by 'That is a lark' could not have been believed on any basis without its being true,

[13] For the latter I have in mind the metaphysical notion of closeness associated with, for example, the treatment of counterfactuals in David Lewis, *Counterfactuals* (New York: Blackwell, 1973).

[14] See Sosa, "Replies," in Greco, ed., *Sosa and His Critics*, pp. 275–325.

[15] Sosa has implemented this general idea in different ways: see Sosa, "Skepticism and Contextualism," *Philosophical Issues*, X (2000): 1–18; Sosa, "Replies," in Greco, ed., *Sosa and His Critics*; and Sosa, "Replies," *Philosophical Issues*, X (2000): 38–42. I have here interpreted his subjunctive conditional (the reason/basis for *p* would be present only if *p*) according to his suggestion in Appendix IV of "Relevant Alternatives, Contextualism Included": "One might hold that in a case where *r* is false, what is required for the truth of '$r \rightarrow p$' interpreted thus is not just that it be true that *p* in the closest worlds in which it is true that *r*, but that it be true that *p* in those worlds and also in other more remote worlds up to a certain (context-determined) threshold."

BASIS is satisfied in both the nearby-imposter and nearby-illusion cases. But we can co-opt the idea behind BASIS by adopting REVISED SAFETY but requiring that counterpart thoughts must have the same—or a sufficiently similar—basis. If we do this, we inherit a version of the dreaded generality problem for method-reliability conditions: What counts as the same basis? Or how similar must the basis be? It would do no good to hold fixed, as Lewis does, every aspect of the subject's "evidential state." Suppose that in the first lark case, any of the imposters would have sounded ever-so-slightly different from any lark. Nevertheless, I am unsafe—and not because of distant, skeptical worlds in which I hear exactly this sound but there is no lark.[16] So the counterpart relation must cover beliefs with slightly different evidential grounds.

It must also cover beliefs with slightly different content. Consider Kripke's case of a region where all the fake barns are yellow, while the one true barn is red. I look out the window, see the red barn, and come to believe that I am driving by a red barn.[17] According to standard safety, this belief is safe, while the (entailed) belief that I am driving by a barn is unsafe. But intuitively my nearby possible false thoughts about yellow barns endanger my actual true thoughts about the red barn. So these are similar enough to be counterparts.[18] But a nearby false thought about how many sheep are on the hillside is not.

It has often been thought that we need both a safety and a reliability condition on knowledge. Reliability alone does not handle the

[16] It is worth noting a further difficulty for Lewis's account. His claim is: *S* knows that *p* iff *S*'s evidence eliminates all not-*p* worlds that are not being properly ignored ("Elusive Knowledge," *Australasian Journal of Philosophy*, LXXIV (1996): 549–67). The notion of evidence here is not understood in terms of what is known by *S*: instead, we hold fixed all of *S*'s experiences, narrowly individuated. But if we individuate the experiences narrowly and the proposition *p* widely, there is a problem. Suppose *S* is looking at an enormous array of pixels, all dark except one ('alpha'), which is bright yellow. *S* looks at alpha and says, 'That is lit'. Moreover, there is a close, salient, experientially indistinguishable world in which the next pixel over (beta) was lit instead. In that world, *p* is false (because pixel alpha is not lit). Of course, in that world *S* does not even believe that *p*; she believes that pixel beta is lit. Nevertheless, Lewis's version of safety counts *S* as not knowing that *p*. This is not a problem for Lewis if he intended that *p* be individuated narrowly, but it is a problem for those of us who take the semantic value of such demonstrative thoughts to be object-dependent. See the related point and excellent discussion in Jonathan Schaffer, "Knowledge, Relevant Alternatives, and Missed Clues," *Analysis*, LXI (2001): 202–08.

[17] Kripke's remarks from which this example is drawn remain unpublished.

[18] Sosa suggests that one's belief that there is a red barn nearby is in some sense *based* on the unsafe belief that there is a barn nearby; whereas in the sailfish case the belief that there is a sailfish nearby is more direct (Sosa, "Replies," in Greco, ed., *Sosa and His Critics*). So in handling Kripke's case it is tempting for a proponent of safety to revise the account so that the necessary condition on knowledge is safety*, where a belief is safe* only if it and any transparent entailments are safe. But the work can also be done by a notion of sufficient similarity of thoughts, which safety theorists must appeal to anyway.

initial lark case properly. My ability to discriminate larks from other birds may be so reliable that there are only five token birds in the world that I would mistake for larks. But if all five of them happen to be in my yard along with a real lark, that is enough to undermine my knowledge. It is the *near* danger of messing up, not the unreliability of my faculty, that is at work in this case.[19] At the same time, standard safety alone does not rule out coming to know a mathematical truth by tossing a coin with the rule: heads, I will accept the Reimann hypothesis; tails, I will deny it.[20] If this method generates a truth, that truth will be necessary, and so there will be no worlds in which I believe that very proposition and it is false. Revised safety, in contrast, can treat nearby false beliefs about different mathematical propositions as counterpart thoughts. It inherits a version of the generality problem, but unlike standard safety, it need not be supplemented by a reliability constraint.

III. REVISED SAFETY AND APRIORITY

Many philosophers are comfortable with fallibilism even when it comes to the phenomenology that typically accompanies a priori knowledge: a thought that to all appearances is a piece of a priori knowledge could turn out to be false. This can occur, for instance, if we are unlucky enough to have flawed internal mechanisms whose deliverances are accompanied with a strong sense of a priori obviousness. But revised safety puts further strain on traditional notions of apriority: for beliefs that would otherwise be paradigmatic pieces of a priori knowledge can be unsafe even when they are true and everything is going well internally.

[19] To insist that my reliable bird-discrimination method becomes unreliable "in this environment" would be to operate with a notion of reliability-in-an-environment that requires infallibility.

[20] DeRose in "Sosa, Safety, Sensitivity, and Skeptical Hypotheses," tries to motivate adding his "strength" condition with this sort of example. Strength is disturbed by nearby worlds in which *S* disbelieves *p* but it is true. But strength cannot handle a case in which I flip a coin to decide whether to judge the hypothesis true—or else to withhold judgment about it. In that case I do not, in any close worlds, disbelieve the hypothesis. One might be tempted to add "counter-safety"—in no close *p* world do I fail to believe *p*—but this is too strong. In many cases of knowledge that *p* one might easily have been distracted and failed to believe *p* (see Sosa, "Replies," in *Philosophical Issues*).

In the belief-withholding coin-toss case, the problem is not that one might have been led to falsehood by the coin-tossing method itself. The problem is that one might have been led to falsehood by whatever procedure caused us to adopt that method. It is the meta-method, if you like, that is the problem. (Suppose the subject is told, by someone she knows to be a great mathematician, that she cannot go wrong using this method; in that case, it may well deliver knowledge.) And of course meta-methods, being themselves methods, can be handled by reliability or revised safety; so DeRose's extra condition is unmotivated.

Consider an example due to John Hawthorne:[21] "A Priori Gas" makes one liable to miscalculate without loss of confidence in one's conclusions. (One becomes disposed to fail to "carry the one," perhaps.) Now suppose I am walking through a dark alley, much of which is filled with A Priori Gas, but I do not actually enter the gas. Had I passed through it at any point, I would have become temporarily disposed to miscalculate, without noticing the change. As things stand, I am lucky enough to avoid it and I have no reason to suspect I am in the vicinity of A Priori Gas. Now suppose that I calculate that 28 + 47 = 75. In a nearby world, I am in a cloud of gas and calculate 28 + 47 = 65, and if we allow that this is a counterpart thought, then by revised safety my actual true belief is unsafe. Therefore I do not know that 28 + 47 = 75, even though everything has gone well internally.[22]

More generally, consider any reliable method that allows us to know things a priori without inference, perhaps by causing us to find propositions primitively compelling. Insofar as a reliable method is fallible, the gas associated with that method can exploit whatever it is that explains those rare occurrences in which the method misfires. It follows that anyone who knows anything by that method will have a possible intrinsic duplicate with the same belief who is near the gas and so unsafe. But these considerations should not lead us to skepticism about the a priori. What matters for knowledge—even a priori knowledge—is whether the subject is actually safe, not whether being internally like the subject guarantees safety.

The social nature of meaning leads to similar results for other paradigmatic pieces of a priori knowledge. I have in mind a semantic picture according to which we all intend to use (even in our thoughts) public terms whose semantic value is settled by community-wide dispositions and patterns of linguistic use. For example, take Jones's belief that every bachelor is eligible. She believes this because she just finds it primitively compelling that bachelors must be eligible. (If challenged, she would say: eligibility is just part of what it is to be a bachelor!) And we can assume that she is right: the Pope is not a bachelor. But suppose that small changes in dispositions and patterns

[21] In conversation; much of the rest of this section grew out of conversation with Hawthorne about issues related to *Something in Mind*, *op. cit.*

[22] Objection 1: A Priori Gas temporarily alters cognitive structure; and close possibilities involving intrinsic changes in the subject should not matter to safety. Reply: This cannot be right in general, lest it make all our beliefs about our own intrinsic states trivially safe. Objection 2: My "method" changes when I pass through a cloud, and thoughts derived from a different method should not be considered counterparts. Reply: If there are "methods" like addition-while-not-failing-to-carry-the-one, what keeps us from ruling out every unsuccessful calculation as the result of a different method?

of use in her community would have lead to a slightly different property being expressed by 'bachelor'. Not everyone finds it primitively compelling to the same degree that bachelors must be eligible; and if a few more people had been disposed to think that noneligibles could be bachelors, the semantic value of 'bachelor' would have been a different property, bachelorhood*, which is compatible with noneligibility.[23] Now, if Jones would still confidently have come to believe the proposition she would have expressed by 'All bachelors are eligible', she would have been wrong. As a result, she is not safe in her actual belief that all bachelors are eligible.[24]

Some will complain that insofar as Jones would not have been a normal user, she would not have counted as really knowing the meaning of the term 'bachelor'. But in the envisaged example, there are no "normal users" of the term, just a range of intuitions about the whether, for someone to count as a bachelor, he must be eligible. Moreover, users with different intuitions may be perfectly interspersed, so that talk of dialects is inapplicable.[25] A further objection: the example, as presented, assumes something like an epistemicist theory of vagueness. Adopting an alternative picture instead would allow us to insist that if 'all bachelors are eligible' is true, very small changes in use of the sort relevant to safety could not make this sentence false—they could only induce vagueness in the predicate 'bachelor' and thus indeterminacy or lack of truth value in the sentence.[26] There is a decision point here: if we count a belief with indeterminate truth value as a failed belief, then Jones does not satisfy revised safety. If not, it will be indeterminate whether Jones knows, because it will be indeterminate whether she is in danger of a false belief. (In the latter case, the force of the example is somewhat mitigated.)

In short, given safety and semantic externalism, it is possible to grasp the necessary connection between bachelorhood and eligibility and yet fail to know that all bachelors are eligible. (Or at best—it may be indeterminate whether one knows it.) Adopting parts of a public

[23] Ignorance due to semantic blindness of this sort is discussed in Williamson, *Vagueness*.

[24] Of course, she is still safe in her belief that all bachelors are male.

[25] This sort of example may lead people to become skeptical about semantics for public language. But it is far from clear that limiting the semantic supervenience base to the use and dispositions of an individual does a better job of avoiding the mess. Given our own indecision and inconsistency, not to mention our apparent reliance on semantic division of labor, the situation does not improve as the base is restricted.

[26] One might say this if one were a supervaluationist, for example. I cannot here consider how the many approaches to vagueness might handle this case. But note the specter of higher-order vagueness; if small enough changes make a sentence indeterminate in truth value, there is pressure to concede that it is actually indeterminate whether the sentence is determinately true.

language into one's mental life is a risky business, because introspection is not a perfect guide to the dispositions and intuitions that constitute the supervenience base for their meaning.

Before considering a final example, let us distinguish two kinds of semantic externalism. A linguistic item or thought is semantically external simpliciter just in case its semantic value does not supervene on the internal features of the speaker (thinker). But an expression or thought is *boldly* external just in case it might lack semantic value altogether in a sufficiently unfriendly environment.[27] In the latter case, whether it has any semantic value, and not just which semantic value it has, depends in part on the subject's environment. Let us assume that names introduced nondescriptively, and used in an ordinary way, are boldly external both in language and in thought. As it happens, names almost always have semantic values in our linguistic community. If I hear friends using the name 'Jane' in conversation, I might come to believe that 'Jane' refers to Jane or even that Jane is Jane, and it is rare for there to be an imminent danger of reference failure. Indeed it seems that following the schematic rule

EMPLOY: If '*N*' is a name in your language and you have no reason to doubt that '*N*' refers, infer that *N* is *N*.

is (for me) a good way to arrive at knowledge. But it would not serve me nearly so well were I unwittingly embedded in a language community with an alarmingly high proportion of empty names. In that case, I might truly but unsafely believe that Jane is Jane or that 'Jane' refers to Jane. For given bold externalism about these thoughts, they would have no content (or gappy content) in cases of reference failure. It follows that I actually know, rather than truly believe, that Jane is Jane in part because of my friendly linguistic environment. (Things do not go better if we fall back to 'If Jane exists, then Jane is Jane'; given bold semantic externalism, this will also express a gappy proposition at best.)[28]

However, the deleterious effects of an unfriendly environment can be counteracted. Suppose that in the case where I pick up 'Jane' from

[27] I am pretending that these items are not individuated by content. More carefully, a linguistic item or thought is boldly external iff a subject tokening such an item has possible intrinsic duplicates that token contentless counterparts of that item.

[28] Granted, one approach treats all atomic gappy propositions as false, and would thus treat the proposition expressed by this conditional as true. (As we have seen, it will also turn out to be the same proposition as that expressed by 'If Vulcan exists, then Bigfoot is Santa Claus'.) But aside from other difficulties with this approach, we have already argued in section 1 that even if some thoughts with gappy content are considered true, they should nevertheless be considered epistemic failures from the safety-theoretic point of view. See footnotes 6 and 7 and the text to which they are appended.

my friends, they had been plotting to fool me with an empty name but then ended up actually talking about a real person by that name. In such a case I am only actually unsafe in my belief that Jane is Jane if I would have been fooled. If certain cues would have made me suspicious enough to avoid adopting the name, then my actual belief is safe. In such a case, let us say that I am *responsive* to the semantic dangers of my environment. Note that one can benefit from one's responsiveness, or from a friendly environment, without realizing it. Suppose I am asked, 'How do you know that *t* is not empty?' Having considered these matters, I may say: 'I have empirical reasons to think that my community is not very unfriendly and also that if '*t*' had been empty, I would have noticed'. If I am worried about the possibility of reference failure, I use this empirical evidence to rule out close content-failure scenarios even before being challenged.

But consider Smith, who unlike me follows EMPLOY without relying on premises involving his friendly environment or his responsiveness. (We will address the issue of *implicit* reliance in a moment.) When he hears Jane being discussed among friends, Smith feels no obstacle in coming to believe such benign-seeming things as that Jane is Jane. He simply has not reflected on the possibility of content failure. This does not mean that Smith is unsafe: indeed, we may suppose that his community is friendly and that he is (unselfconsciously) responsive to most of the danger that remains. Likewise, we may suppose that Smith finds it primitively compelling that all bachelors must be eligible, and unlike Jones he is not in danger of semantic shift. Can Smith reap the epistemic benefits of exploiting an external source of reliability without involving it in (even the tacit) justification of his beliefs? Can he enjoy this knowledge, blissfully ignorant of his semantic good fortune? As far as revised safety is concerned, yes; for to be safe does not require *knowing* that one is safe.[29]

If we grant that Smith knows, we face a second question: Should we count his knowledge as a priori? It might be argued that a priori knowledge cannot be environment-dependent in the way just described; but as we have seen, that would rule out even mathematical knowledge based on simple calculations. To disallow Smith's knowledge as an instance of the a priori, it might be argued that unlike mathematical knowledge, it requires some sort of implicit justification by empirical premises. Perhaps one cannot know that Jane

[29] But perhaps in order to *know* that one knows, one must know that one is safe. In section IV, I discuss some of the implications of revised safety on iterated knowledge.

is Jane or that bachelors are eligible unless one is somehow prepared to provide some a posteriori evidence to justify the belief if challenged with the possibility of reference failure or semantic shift, respectively.

This raises delicate issues involving the nature of implicit justification. For if someone challenges my belief that 28 + 47 = 75 on the grounds that I might be in A Priori Gas (or drunk) and therefore miscalculating without noticing it, I may trot out some empirical evidence that I am not in A Priori Gas (or drunk). Does this mean that I am always somehow implicitly relying on this evidence for my mathematical beliefs, or at least for my second-order belief that I *know* the mathematical truth?[30] One could try to sever the tie between answers to challenges and implicit justification, maintaining that Smith's beliefs are somehow implicitly justified by empirical evidence, while my mathematical belief is not—even though we both may appeal to such evidence when challenged. But I doubt this distinction could be achieved with a natural conception of implicit justification. For evidence about the reliability of our calculating skills is available to us all. Indeed, it is difficult to imagine lacking such evidence, but if we did I suspect we would not be so confident in the deliverances of our calculations. Does such evidence always play an implicit justificatory role in our beliefs? I have no good answer to this question. My point in this section has been to illustrate with revised safety that these cases do not fit easily within the traditional distinction between a priori and a posteriori knowledge.[31]

IV. REVISED SAFETY AND SELF-KNOWLEDGE

Given revised safety and bold externalism for demonstrative thought, we can easily demonstrate that "strong privileged access" is false:

(SPA): Necessarily, if *S* is thinking that *p*, then *S* is in a position to know a priori that she is thinking that *p*.[32]

[30] Suppose it does. One could argue that it is nevertheless *possible* for the mathematical belief (and even the second-order belief) to be justified without reliance on empirical data; whereas the belief that Jane is Jane could *never* be adequately justified without such reliance (except perhaps if the believer is Jane herself). Then, if one uses 'a priori' to mean 'not explicitly reliant on empirical justification, and in principle capable of being justified without even implicit reliance on empirical justification', one could insist that my mathematical beliefs are priori and Smith's beliefs about Jane are not.

[31] I suspect it is vague whether the pieces of knowledge in question are a priori. Terms of art are not after all immune to vagueness, especially those with so long and varied a history as 'a priori'. I have no objection to putting 'a priori' to a precise use, as long as one acknowledges the technical nature of the term.

[32] From Michael McKinsey, "Forms of Externalism and Privileged Access," *Philosophical Perspectives*, XVI (2002): 199–224. His version reads "*x* can in principle know a priori."

Happily, this can be done without relying on reductios of the sort that have been brought to light by Michael McKinsey and Paul Boghossian and have been discussed at great length in the literature on externalism and privileged access.[33] And it can be done without worrying about whether introspective knowledge counts as a priori, because we can show that some thinkers that *p* are not in a position to know—a priori or otherwise—that they are thinking that *p*.

Recall my demonstrative lark-thought in the case where I am actually hearing a lark but I have also been hallucinating lately. We saw in section I that this belief is not safe. Now suppose that by introspection, I come to form a thought I would express with 'I'm thinking that that is a lark'. Note that in ascribing a demonstrative belief to me, this sentence itself *uses* and does not *mention* the demonstrative. What would it express in the absence of a referent? If we hold, with semantic orthodoxy, that demonstrative expressions are boldly external even within the scope of propositional attitude ascriptions,[34] then this sentence would have failed to have any (complete) content. And again, given bold externalism about the corresponding thought episode, it would likewise not have had any (complete) content. It follows from this along with revised safety that if such a case is nearby, I do not know that I am thinking that that is a lark.[35]

Unlike the standard arguments for the incompatibility of semantic externalism and SPA, the argument from safety does not concern the question whether it is possible to deduce facts about the world from a piece of introspection. In fact, our subject does not go through any process of reasoning at all. As a result, the most common ways to rebut McKinsey's and Boghossian's arguments have no application to the argument from revised safety.

We have denied knowledge to subjects whose introspective beliefs are unsafe, but what should be said about those of us who practice

[33] See McKinsey, "Anti-Individualism and Privileged Access," *Analysis*, LI (1991): 9–16; and Boghossian, "What the Externalist Can Know A Priori," in Wright, Smith, and Macdonald, eds., *Knowing Our Own Minds* (New York: Oxford, 1998), pp. 271–84.

[34] I am not assuming that propositional attitude contexts are "Shakespearian": namely, that co-referential singular terms can be substituted salva veritate in those contexts. I am assuming that, whatever else this belief-ascription says, it must relate me to a singular proposition involving a specific lark or fail to have any complete content.

[35] Brown considers the idea that the epistemic status of this sort of belief can be undermined by a close world in which I falsely believe something nonsingular, such as the proposition I would express by 'I am having a thought about some lark or other'. See Brown, *Anti-Individualism and Knowledge* (Cambridge: MIT, 2004). But given revised safety, the epistemic status of the singular introspective thought is directly undermined by the fact that in a close world, I token a thought internally just like it but lacking in content.

safe introspection? Revised safety and object-dependence are compatible with a modified privileged access thesis, according to which if a subject knows that *p*, she is typically in a position to know that she knows that *p*. So insofar as we take ourselves to know by introspection the content of many of our boldly external beliefs, we do face a self-knowledge puzzle after all.

Here is a version of the problem using our terms.[36] Suppose again that I have adopted the name 'Jane' after having heard much talk about a person by that name. This time I am in a friendly environment, as well as tacitly responsive to the name-producing reliability of my community, like Smith in section III. Then I can know by introspection

(1) I am thinking that Jane is great.

Now, if I am aware that proper names are boldly external, and I know that I intend to use 'Jane' as a proper name whose referent is settled by my interlocutors, I am in a position to conclude that Jane's existence is a necessary condition for contentful Jane-thoughts.

That is, I can know

(2) If I am thinking that Jane is great, then Jane exists.

This coupled with knowledge of (1) appears to put me in a position to know that Jane exists. Set aside whether we ought to count the premises as genuinely priori; it is puzzling enough that I am in a position conclude that Jane exists from an introspective premise along with a premise about a general feature of the semantics of my language.

[36] The puzzle that follows is more similar to that of Boghossian, "What the Externalist Can Know A Priori," than it is to the one originally set forth in McKinsey, "Anti-Individualism and Privileged Access." McKinsey argues that from the truth of (a certain kind of) semantic externalism we can conclude that 'The proposition that Oscar is thinking that water is wet logically implies' a proposition involving the existence of objects external to Oscar. (The emphasis here is on 'logically implies'.) But then, according to McKinsey, if Oscar can know a priori that he is thinking that water is wet, 'Oscar can just *deduce* E [the proposition about external objects] from something he knows a priori, and so he can know E itself a priori'. In other words, Oscar himself need not know the truth of externalism; it just follows from externalism that he can deduce E from the introspective knowledge. McKinsey reiterates the point in his "Transmission of Warrant and Closure of Apriority," in Susana Nuccetelli, ed., *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge: MIT, 2003), pp. 97–116, on p. 98.

I find this version of the puzzle less compelling because it puts so much weight on a notion of "logical implication" between propositions (not sentence types), a relation that McKinsey says goes beyond metaphysical implication. McKinsey requires that externalism commits us to the following thesis: the proposition that Oscar is thinking that water is wet transparently implies some proposition about the existence of water. (Where *p transparently implies q* just in case, if anyone knows that *p*, that person is in a position to simply deduce *q* from *p*.) But this thesis needs more support than he provides.

The puzzle should have a familiar ring to epistemologists: it parallels Fred Dretske's classic closure puzzle. Sandra, a normal zoo-goer, knows that the animal before her is a zebra. So is she in a position to deduce that it is not a mule cleverly disguised as a zebra? (She knows the steps that would take her from one to the other.) Or again, if you know that you will be going to work tomorrow, does this put you in a position to know that you will not win the "great surprise heart-attack lottery" tonight?[37] The self-knowledge puzzle is just a particular kind of closure puzzle in which the subject comes to know something that he should not be in a position to know by way of introspection plus general semantic knowledge. A range of responses is available for any closure puzzle:[38]

(a) Revoke the original knowledge. The skeptical position.
(b) Embrace the reductio. The anti-skeptical conclusion.
(c) Deny closure.
(d) Adopt contextualism or "sensitive invariantism."

I will set aside (a) and (b). Skepticism is a last resort, and I do not want to allow that a good way of coming to know that I will not win the lottery (or die) tomorrow is by knowing that I am going on a modest vacation in a week. That leaves (c) and (d).

In the case of the self-knowledge puzzle, we have two further options:

(e) Deny bold semantic externalism.
(f) Exploit gappy propositions to reject the move from (1) to (2).

Much of the literature on externalism and self-knowledge has focused on option (e), which as I will argue below is a red herring. Moreover, while (e) has some plausibility when it comes to predicates like 'water',[39] which figured in the original McKinsey example, it is less

[37] For these and related cases, see Dretske, "Epistemic Operators," this JOURNAL, LXVII, 24 (December 24, 1970): 1007–23; Cohen, "How to Be a Fallibilist," *Philosophical Perspectives*, II (1988): 91–123; Jonathan Vogel, "Are There Counterexamples to the Closure Principle?" in Michael D. Roth and Glenn Ross, eds., *Doubting: Contemporary Perspectives on Skepticism* (Boston: Kluwer, 1990), pp. 13–27; DeRose, "Knowledge, Assertion, and Lotteries," *Australasian Journal of Philosophy*, LXXIV (1996): 568–80; Hawthorne, *Knowledge and Lotteries* (New York: Oxford, 2004).

[38] Much of the recent literature on McKinsey's puzzle focuses on the issue of "warrant transmission." But note that the stance according to which the subject's warrant fails to transfer to the conclusion is not a further option to be added to (a)–(f).

[39] The idea would be that even if we have always been embedded in the worst sort of environment (such as Boghossian's "Dry Earth"), 'water' still has had some semantic value or other—for instance, the functional property *being clear, drinkable, liquid, and so on*. See Boghossian, "What the Externalist Can Know A Priori." (Moreover, while having a belief about that property entails the *existence* of that property, it does not entail that

plausible for names and demonstratives. A neglected alternative is (f): bold externalism only requires that 'Jane' would have no semantic value in the absence of a referent. This is consistent with 'Jane exists' having gappy content. We might then hold quite naturally that an attitude ascription of the form '*N* thinks that *S*', where '*S*' is replaced by a sentence with gappy content, is true just in case the individual referred to by '*N*' is belief-related to the gappy proposition expressed by '*S*'. If this is the case, then knowledge of general semantic facts does not put one in a position to infer (2) from (1).[40]

However, I will set (e) and (f) aside because semantic externalism is something of a red herring in this vicinity. To see this, consider that one can cook up closure puzzles involving knowledge of one's own mental states that have nothing to do with semantic externalism. Consider my second order belief in the mathematical case:

(3) I know that 28 + 47 = 75.

If I know (3), and also that knowledge must be safe, this seems to put me in a position to deduce

(4) There is no A Priori Gas around.[41]

This self-knowledge closure puzzle has just as much force as the original one; but (e) and (f) do not apply here. Neither, of course, are they available for standard closure puzzles that do not involve self-knowledge. It would be best to treat all closure puzzles in the same fashion, so that (to mix metaphors) we can kill many birds by biting one bullet. And that is exactly what revised safety allows us to do, as I hope to show.

Let us return to the option of denying

CLOSURE: If *x* knows that *p* and *x* properly deduces *q* from *p*, then *x* knows that *q*.

---

there are any *instances* of it.) I will not contest this approach, except to point out that it seems much less plausible for proper names, demonstratives, and the like.

[40] Things are more complex if in the example I know the second-order proposition expressed by 'I know that I am thinking that Jane is great'. Then, if I know that 'Jane' is boldly external and that revised safety does not allow knowledge of gappy propositions, it would appear that I can infer (2). However, it might be suggested that even if 'Jane' is empty, the proposition expressed by 'I am thinking that Jane is great' should not count as gappy, at least for the purposes of revised safety. For one could hold that the emptiness of 'Jane' does not hinder 'that Jane is great' from functioning as a complete term for a proposition, so in the relevant sense the belief ascription has a complete content.

[41] Thanks to Hawthorne here.

A few explanations have been offered for why closure might fail in these cases, but I will focus on the appeal to sensitivity as a necessary condition on knowledge:

SENSITIVITY: Had *p* been false, *S* would not have believed *p*.[42]

If a piece of knowledge must be sensitive, it follows that even when a subject performs a proper deduction from a single known premise, the subject may not be in a position to know the conclusion. For instance, Sandra's first belief about the zebra is sensitive, while the second is not.

How can sensitivity help when it comes to the self-knowledge puzzle? The "bad" conclusion is that Jane exists. But if Jane had not existed, I would not have believed that Jane existed. For one thing, I would not have heard anyone using 'Jane'. For another, given bold externalism, there would be no proposition about Jane to express, Jane being unavailable to serve as a component of it. So there are no worlds at all, let alone a close world, where Jane does not exist but I believe that Jane exists. For those worlds of evaluation *at* which the proposition is false are worlds *in* which it is not available to be the content of any thought.[43] So sensitivity is no use.

In fact, if sentences containing names and demonstratives typically have object-dependent contents, sensitivity cannot even do the work it was originally designed for. Consider a slight modification of our original example with Sandra. This time she uses a demonstrative in her thought: she believes that *that* is a zebra and then concludes that *that* is not a cleverly disguised mule. Assuming a plausible thesis of species-essentialism, both of these beliefs are trivially sensitive because they involve necessary propositions.[44] This point is not a counterexample to sensitivity as a necessary condition on knowledge, but it is devastating for sensitivity nonetheless. For the condition was *introduced* to block Sandra's inference from 'the animal is a zebra' to 'the animal is not a cleverly disguised mule'. How can the condition be taken seriously if it does not also block the inference from 'that animal is a

[42] Proponents of sensitivity and its variants include Dretske, "Conclusive Reasons," *Australasian Journal of Philosophy,* XLIX (1971): 1–22; Robert Nozick, "Knowledge and Skepticism," in his *Philosophical Explanations* (Cambridge: Harvard, 1981), pp. 167–290; and DeRose, "Solving the Skeptical Problem."

[43] Alternatively, if we interpret the antecedent of SENSITIVITY as asking us to look at worlds *in* which *p* is false—rather than *evaluated at which p* is false—then the condition will be trivially satisfied because if *p* exists it cannot be false.

[44] A proponent of sensitivity might suggest a theory of counterfactuals for which some (but not all) counterfactuals with impossible antecedents are false. Even if such a theory could be made plausible and could give the right result in this case, it would only help with the demonstrative zebra case, not with the introspective closure case.

zebra' to 'that animal is not a cleverly disguised mule'? Perhaps sensitivity can be revised (in the spirit of our revision of safety) to take care of these problems, but at the moment I do not see exactly how this could be done.[45]

We are left with (d). The basic contextualist answer is straightforward. Two speakers may utter a sentence of the form *S knows that p at t* about the very same *S*, *p*, and *t*; and yet one speaks the truth and the other speaks falsely. In particular, while the epistemic credentials of a subject's belief might pass muster in one context, they might not be good enough when the possibility of error becomes salient. So, while in ordinary contexts 'Sandra knows there is a zebra in the cage' is true, it is not true in a context where it is particularly salient that Sandra might have messed up. In such a context, a different relation is expressed by 'knows', one that Sandra does not bear to *either* the proposition that there is a zebra in the cage, or the proposition that there is not a cleverly disguised mule in the cage. The result is that, as Stewart Cohen puts it, "if we evaluate the closure principle relative to a fixed context, thereby fixing the standard, it comes out true."[46] So the letter of the closure principle, at least, can be observed.

Unfortunately, contextualism is usually cashed out in a way that is not particularly amenable to the semantic picture we have been working with. For suppose we flesh out the intuition about "salient chances of error" in terms of contextually-sensitive thresholds on the degree of safety required for knowledge, as Sosa, Keith DeRose, and Mark Heller do:[47]

> THRESHOLD: For '*S* knows that *p*' to be true in a context *C*, *S*'s belief that *p* must be safe to a degree specified by *C*.

Given a gradable version of standard safety, a belief is safer the more distant are the closest worlds where *S* falsely believes that *p*. While this

[45] For other problems with sensitivity, see Sosa, "Skepticism and Contextualism"; Vogel, "Tracking, Closure, and Inductive Knowledge," in Steven Luper-Foy, ed., *The Possibility of Knowledge* (Lanham, MD: Rowman and Littlefield, 1987), pp. 197–215; and Schiffer, "Contextualist Solutions to Scepticism," *Proceedings of the Aristotelian Society*, XCVI (1996): 317–33. Just to mention two: (1) a belief of the form 'I am not wrong in thinking that *p*' cannot be sensitive; and (2) though my belief that I am not a brain in a vat is insensitive and therefore not knowledge (a result that sensitivity theorists like), my belief that I am not a brain in a vat with no auditory sensations is also insensitive and therefore not knowledge (a result that no one should like).

[46] Cohen, "Contextualism and Skepticism."

[47] DeRose, "Solving the Skeptical Problem," and "Sosa, Safety, Sensitivity, and Skeptical Hypotheses"; Sosa, "Skepticism and Contextualism," and "How to Defeat Opposition to Moore," *Philosophical Perspectives*, XIII (1999): 141–53; Heller, "The Proper Role for Contextualism in an Anti-Luck Epistemology."

may get the right result for the standard zebra case, the explanation for why Sandra does not know the conclusion does not apply to the demonstrative zebra case, because in that case both beliefs will be maximally safe. Neither can the explanation apply to the introspective closure puzzle: my belief that Jane exists is also maximally safe, because in worlds where Jane does not exist I could not believe that she exists.[48]

Here again revised safety is in its element:

> GRADABLE: A belief is safer the more distant are the closest worlds in which a counterpart fails.

Given a gradable version of revised safety, and applying THRESHOLD, we get the conclusion we want. Take the demonstrative zebra case. When ascribers are properly ignoring possibilities with cleverly disguised mules in them, they can truly utter, 'Sandra knows that that is a zebra'. But the mere act of considering Sandra's chain of reasoning to the conclusion that *that* is not a cleverly disguised mule makes it impossible to ignore possibilities in which there is a cleverly disguised mule in front of her instead. And in those worlds, Sandra has a false counterpart belief—in a different proposition. The context has shifted, and now neither demonstrative proposition falls under the extension of 'knowledge'. Closure, relativized to contexts, holds firm.

Revised context-dependent safety also handles our self-knowledge closure puzzles. In normal contexts, we may ascribe a priori knowledge to someone who thinks that Jane is Jane. But when we consider the reasoning that leads to (2)—reasoning that involves semantic intuitions about the meaning of empty names and so on—we can no longer ignore the possibility that Jane has picked up an empty name. This shifts our context and requires safety to extend out to such possibilities as well, which it does not. So when we evaluate the status of the resulting belief that Jane *exists*, we are in a context where neither (1) or (2) count as known. A similar story can clearly be told about my knowledge of (3) and (4).[49]

[48] The same thing is true of Lewis's approach, which tells us that *S* knows that *p* iff *S*'s evidence eliminates all not-*p* worlds that are not being properly ignored, and that different worlds are properly ignored in different contexts (Lewis, "Elusive Knowledge"). Assuming object-dependence for demonstratives, this does not explain why we cannot count Sandra as knowing that *that* is not a cleverly disguised mule on the basis of her knowledge that *that* is a zebra. For there are no not-*p* worlds to be ignored in either case.

[49] For an application of contextualism to introspection puzzles, see Jakob Hohwy, "Privileged Self-Knowledge and Externalism: A Contextualist Approach," *Pacific*

A related approach to closure puzzles is "sensitive invariantism."[50] According to this view, 'knows' does not express different relations in different speaker contexts; but whether *S* knows that *p* depends to a surprising degree on the *S*'s own context. For example, high stakes relevant to the belief and possibilities of error that are salient to the *subject*, can undermine knowledge. This is not a variety of contextualism because 'knowledge' has but one semantic value in any context—it is just a relation that is at the mercy of practical features of the subject's environment. The zebra closure puzzle is handled by saying that Sandra's knowledge of the original zebra-proposition dissipates as she starts to consider the possibility that the zebra is a cleverly disguised mule. But closure need not be denied, as long as we are careful to understand it as the thesis that if *S* knows that *p* and *S* deduces *q* from *p while maintaining her knowledge of p*, then *S* knows that *q*.

Once again the view needs to be tweaked in order to handle semantic externalism. Hawthorne uses examples where subjects undermine their knowledge that *p* by considering indistinguishable possibilities in which *p* is false. But we must be careful not to limit the sorts of nearby and salient possibilities that destroy knowledge to possibilities of *false belief that p*. Otherwise, for reasons that should now be familiar, neither the demonstrative zebra case nor the self-knowledge puzzle could be properly resolved.

Before concluding this discussion of self-knowledge, it is worth stressing that a subject can be safe in her self-knowledge even if she could easily have undergone an unnoticed semantic shift—as long as her thought would not have failed. Safety thus does not require that, if *S* knows that *S* is thinking that *p*, then *S* must be in a position to discriminate the thought that *p* from any nearby thoughts with different content. This is the right result; following Burge and others, I consider this sort of discrimination principle too strong.[51] The ability to discriminate the proposition that one actually believes from other propositions is relevant to whether one counts as "knowing which"

Philosophical Quarterly, LXXXIII (2002): 235–52. Hohwy employs the contextualism of Lewis, "Elusive Knowledge," and so is subject to the objections of footnotes 16 and 48 above. In particular, assuming bold externalism, his approach cannot handle any closure puzzles involving names and demonstratives, including the demonstrative zebra case and the case involving (1) and (2).

[50] See Hawthorne, *Knowledge and Lotteries*.

[51] See footnote 17 above; see also Burge, "Individualism and Self-Knowledge," this JOURNAL, LXXXV, 11 (November 1988): 649–65; Kevin Falvey and Joseph Owens, "Externalism, Self-Knowledge, and Skepticism," *Philosophical Review*, CIII (1994): 107–37; Anthony Brueckner, "Ambiguity and Knowledge of Content," *Analysis*, LX (2000): 257–60.

proposition one believes; but judgments of knowing which, like judgments of knowing who, are sensitive to an entirely different set of contextual parameters than are judgments of knowing that.[52]

To sum up this section: the self-knowledge puzzle should be treated as any other closure puzzle. But sensitivity is of no use when it comes to closure puzzles of self-knowledge, and runs into trouble even with standard closure puzzles involving demonstratives. Fortunately, pairing revised safety with either contextualism or sensitive invariantism allows us to resolve all of our closure puzzles in a single stroke.

DAVID MANLEY

University of Southern California

[52] This contrast is worthy of more discussion than I can give it here. But note that one can know a man but still not count (at least in certain contexts) as knowing who he is. For example, you may get to know your next door neighbor quite well and not realize he is the son of Little Richard; in certain contexts where his paternity matters, you will not count as knowing who he is. In analogous cases, one can know a proposition but still not count (in certain contexts) as knowing which proposition it is. For more on the factors that govern the felicity of "knowing who/which" claims, see Steven E. Boër and William G. Lycan, *Knowing Who* (Cambridge: MIT, 1986).