

8
Knowing that One Knows What One Is Talking About
Susana Nuccetelli

Twin-earth thought experiments, standardly construed, support the externalist doctrine that the content of propositional attitudes involving natural-kind terms supervenes upon properties external to those who entertain them. But this doctrine in conjunction with a common view of self-knowledge might have the intolerable consequence that substantial propositions concerning the environment could be knowable a priori. Since both doctrines, externalism and privileged self-knowledge, appear independently plausible, there is then a paradox facing the attempt to hold them concurrently. I shall argue, however, that externalist claims about the dependence of content on environmental factors presuppose certain theses about the semantics of natural-kind terms that, if sound, would make those claims eligible for empirical justification instead. In fact, that is the only interpretation of their epistemic status that could square with the standard conclusion from twin-earth cases. Furthermore, it can be shown to solve the paradox of externalism and self-knowledge in a more doxastically conservative way--accommodating precisely each of the well-accepted intuitions about knowledge, closure and the individuation of content given up by available competitors.

I

Self-ascriptions of propositional-attitude contents strike us as being epistemically privileged when contrasted, e.g., with beliefs about the physical world. It is only in the case of the former that it makes sense to vindicate special epistemic access, since it seems *prima facie* implausible that, under normal circumstances, the justification of those beliefs requires either

inference or perceptual evidence. In addition, there is the presumption that, from the first person perspective, under normal circumstances, beliefs of that sort are truth-warranted. And, were experience to be construed narrowly as *sense* experience and knowledge independent from the latter as a priori, although beliefs about the physical world would not qualify for a priori knowledge understood in this way, those about one's own propositional-attitude contents (which appear generally true and justified independently from sense experience) would.

But once this view of a priori knowledge is accepted, then externalism, a theory entirely supported by philosophical argument and thought experiments, turns out to be a priori too--as do equally beliefs deduced from that theory and other a priori premises. And now a *reductio* seems to stand in the way of holding both externalism and privileged self-knowledge. For then one could come to know entirely a priori (by simple deduction) that there are, for example, certain natural kinds in the environment. But that clearly amounts to a substantial matter of fact which could be known, if at all, only a posteriori. As I have argued elsewhere (Nuccetelli 1999, 2001), the epistemic status of beliefs about the empirical world is in no way analogous to that of those about one's own current thought contents, simply because to invoke first-person authority for the former scarcely makes any sense--as it would amount to the claim that, under normal circumstances, such beliefs are truth-warranted. Furthermore, unlike self-ascriptions of content, no belief about the existence of a certain natural kind can be directly justified, since, first, whether any empirical belief qualifies for justification of that sort at all is a matter of dispute. Secondly, suppose one were to assume some version of direct realism (which for the problem of concern here need not be assumed in any case), then although this might entail holding that an individual could come to know directly that, for example, there is a green patch or even a tree before him, it would nonetheless be odd to construe direct realism as claiming direct knowledge--

i.e., knowledge independent of either evidence or inference--of propositions such as that *beech trees exist*.

The argument so far outlined generates at least a paradox for the attempt to hold both externalism and privileged self-knowledge, for these doctrines each seem as well supported individually as is the claim that they are incompatible. Underwriting the argument for their incompatibility is the principle that a priori justification and knowledge are closed under known entailment, which is weaker than other forms of closure and therefore more plausible (even when, as in the case of other rules of reasoning, it appears difficult to justify in a non-circular way¹). Be that as it may, instances of reasoning that mirror this principle seem intuitively acceptable, and can thus be taken to support it--as may those that flout it and result only in flagrant inconsistencies. The incompatibilist's *reductio* trades on the latter. For the attempt to hold both externalism and privileged self-knowledge seems to flout closure under a priori knowledge, sanctioning flagrant inconsistencies of this sort:

- 1 Oscar can know a priori that his current propositional attitude involves a natural-kind term *k*.
- 2 Oscar can know a priori that, if his current propositional attitude involves a natural-kind term *k*, then there is a certain natural kind *K* in his environment.
- 3 Oscar could not know a priori that there is a certain natural kind *K* in his environment.

Given (1) and (2), supported by independently plausible intuitions about self-knowledge and the individuation of mental content respectively, closure prescribes the rejection of (3)--a claim that rests on equally plausible intuitions about knowledge of the empirical world. To restore consistency, one of these must go, but which one? Hardly (1) or (2), since as we saw, self-

ascriptions of propositional-attitude contents appear every bit as a priori justified as the claim that those contents may often depend in part upon factors in the physical environment. When held concurrently, however, (1) and (2) have the absurd consequence that one could come to know entirely a priori substantial propositions about the physical world, just by reflecting on the contents of one's own words and thoughts and their externalistic entailments--and this conflicts with (3).

A resolution of this paradox appears to require the rejection of either privileged self-knowledge, externalism, or closure. Otherwise we are left with the absurdity of holding that one can know a priori substantial propositions about the environment. Yet each of these three is independently plausible. Thus, even though some such response might provide a venue for resolving a paradox that may be overdetermined in any case, each would score poorly in doxastic conservatism. But a closer look at the semantic commitments of externalism may provide a strategy to resolve it with a good share of that epistemic virtue.

II

Twin-earth externalists (hereafter, 'externalists') are committed to certain theses about the semantics of natural-kind terms, which can in turn be shown crucial to determining the epistemic status of externalist entailments from propositional-attitude contents to environmental conditions. First, given twin-earth cases, the meaning of genuine terms of that sort would depend in part upon their reference, but then it seems to follow that the content of propositional attitudes involving any such term would likewise be partially determined by the reference of that term. Standard twin earth cases, however, also require that in any possible worlds where tokens of natural-kind terms refer at all, they necessarily pick out samples of exactly the same substances

or species referred to by such terms in the world where their extension was initially grounded. Recall twin earthian T-Oscar, the exact replica of earthian Oscar in all his internal properties (non-intentionally described), who often utters the sound 'water' to refer to a substance superficially identical to water in Oscar's world but having an inner composition that differs radically from H₂O. Here the externalist urges that T-Oscar neither possesses the natural-kind term 'water,' nor could have propositional attitudes whose content might be intentionally described as involving that concept.² Yet such a conclusion would fail to follow unless externalism were taken to rest upon the semantic theses outlined above.

In fact, twin-earth thought experiments were first proposed as arguments to undermine a certain semantic account, traditional Fregeanism, thought to fuel the conclusion that in spite of radical differences in the reference of some natural-kind term, internal replicas such as Oscar and T-Oscar could entertain propositional attitudes involving exactly the same content. This would indeed follow when Fregeanism is construed as holding, first, that all components of propositions, whether singular terms or general ones, have meanings that supervene entirely upon the local properties of those who entertain them in their speech or thought; and secondly, that meanings of that sort (hereafter, 'Fregean senses') completely determine the extension of any such component. But then chemically ignorant, internal replicas could certainly entertain propositional attitudes involving exactly the same natural-kind term in scenarios where tokens of their term picked out substances of altogether different inner constitutions. Since traditional Fregeanism cashes out the senses of natural-kind terms as the speaker's non-demonstrative ways of thinking about certain properties, and takes senses of that sort to determine the extension of those terms (the referent property), therefore if internal replicas came to associate exactly the same senses with superficially identical substances or species, it would then be possible that

those replicas should have words and thoughts involving exactly the same natural-kind terms in worlds where the inner properties of the substances or species referred to by tokens of their terms differed radically. If traditional Fregeans must countenance such scenarios (as their theory is usually construed—e.g., Searle 1983), then it is difficult to see how they could accommodate the externalists' doctrine.³

For one thing, it follows from standard externalist thought experiments that, as in the case of the individual objects referred to by logically proper names, the properties picked out by tokens of certain predicates must also be the same in any possible world where those terms have an extension at all (which, as argued by Recanati 1993, does not entail holding that they are purely referential). And crucial to the externalist conclusion from twin-earth examples is the metaphysical claim that, in possible worlds where natural-kind terms have an extension, they necessarily pick out samples of natural kinds that have the same inner composition as the exemplars referred to by those terms in the world where their extension was initially grounded. Of course, given externalism, for any natural kind to fall within the extension of a natural-kind term conventionally available in a linguistic community, there must be a referential link between the term and some substance or species, originally established by causal commerce of the community's speakers with the referent natural kind and later transmitted to others through social interaction. Furthermore, it is not their Fregean senses, but their reference instead, fixed by causal relations of speakers with natural kinds, that individuates natural-kind terms, together with the content of any proposition embedded in a psychological attitude containing terms of that sort.

Note that these semantic intuitions invite a certain view about the epistemic status of propositions in cases where they appear to involve natural-kind terms. For if those intuitions are correct, then whether the content of any such proposition actually involves genuine terms of that

type would depend upon whether the individual who entertains the proposition in his speech or thought has in fact entered some causal chains relevant to fixing the reference of occurring predicates into certain substances and species in the environment (either by his direct acquaintance with these, or by interaction with other members of the linguistic community whose referential use of such predicates traces back to those of speakers who have introduced them as a result of direct acquaintance). One's belief that any of these relations obtain, when justified and true, would of course be eligible to count as knowledge in some sense, but merely of the empirical sort. Imagine a world that differs from ours only in that tokens of a certain word, 'water,' have always been empty there--like our 'phlogiston,' no causal chain ever related that word, conventionally available in the linguistic community, with some natural kind. By hypothesis, then, native speakers of that world are part of no causal chain grounding the extension of that term in an external property. In such a scenario, whatever the contents of a native's sincere assertion containing that term might be, they could neither involve *the property of being H₂O* nor be intentionally characterized as expressing, for example, the belief *that water is wet*.

But there is no need to appeal to imagination here, since memory, when applied to Western science (whose history of ontological shrinking is also one of semantic elimination) will do just as well. Recall, e.g., the notoriously bogus scientific term 'luminiferous ether,' taken to pick out the property of being a certain radiation-transmitting medium filling all unoccupied space. No doubt physicists at some point entertained propositional attitudes involving this putative natural-kind term, sincerely thinking that they were referring to a medium of that sort. (Their term, however, regularly picked out nothing external at all, and was finally discarded shortly after the failed Michelson-Morley experiment of 1887.) In such cases, intentional

description would of course fall short of involving a genuine natural-kind term individuated by environmental conditions.

When it comes to putative substance- and species-words, the possibility of extensional emptiness seems to suggest that, in order to be knowledgeable, intentional description requires investigation of the causal history of those words. Tokens of a declarative sentence (type) involving such words, in scenarios where they do have referential links to substances or species in the environment, would admit intentional characterization along externalist lines--but certain propositional-attitude contents may qualify only for internalist intentional description in situations where all embedded terms supervene entirely upon the local properties of individuals.

Charitably construed, externalism is compatible with this conclusion. At the same time, precisely because externalists may countenance the supervenience of some such contents upon local properties of individuals, they seem committed to hold that the problem of determining whether any propositional-attitude contents depend entirely upon local properties of individuals, or partially upon non-local ones, amounts to an epistemic question that cannot be settled just by thinking. Clearly, externalist entailments from propositional-attitude contents to the environment, if they obtain, count as metaphysically necessary. But once the scope of externalism is properly understood, together with its semantic commitments, those entailments seem to presuppose propositions concerning the existence of referential links between terms, on the one hand, and substances and species in the speaker's environment on the other. Belief in propositions of that kind, when justified and true, constitutes knowledge in some sense, but only of the empirical sort.⁴ That knowledge of content is a posteriori, then, appears the only view compatible with the externalist's standard conclusion from twin-earth thought experiments. At the same time, this

view can solve the paradox facing externalism and privileged self-knowledge, provided externalists can support their semantic intuitions, to which I now turn.

III

For most of the twentieth century, philosophical semantics, chiefly under the influence of Frege (1952) and some of Russell's writings (1905, 1918-19), offered an analysis of the semantic contribution of certain terms to the propositions in which they occur that was not to be seriously challenged until near the end of that century. Within that analysis, tokens of 'Laika is a dog' would be taken to express a proposition ascribing a certain property (in this case, being a dog) to whomever fits some definite description (e.g., the first Russian experimental animal sent into space). New theorists of reference--whose theory is of course not new at all, having been sketched by Mill, in his *System of Logic*, in 1843--famously objected that the semantic contribution of a name could be construed as that of a definite description (or a cluster of them) and thus took the Frege-Russell analysis to misrepresent the form of the expressed proposition. On the new theory (Kripke 1972, Kaplan 1977), that proposition must instead be cashed out as a singular, Russellian one, which predicates a certain property of Laika, an individual object that is a constituent of the proposition.

Although externalism typically focuses upon ascriptions of *de dicto* content in cases where the propositions embedded in psychological attitudes contain natural-kind terms, much could be learned about the epistemic status of such ascriptions by reflecting upon the analogous case of belief involving singular propositions. For, assuming that there are any of the latter, then tokens of a certain sentence type could express a proposition of that kind provided that the sentence contained at least one term capable of picking out a certain individual entity--that is,

provided the expressed proposition has at least one singular term functioning as a logically proper name capable of contributing to it the referent object itself. But the actual semantic contribution of a singular term to the proposition in which it occurs crucially depends upon the causal history of that term. For to invoke its referent, a term must latch onto it as a result of some causal commerce of its users. On the causal account of reference often favored by direct theorists, speakers must have had causal contact with the referent object itself, or with other speakers whose use of the term may be traced back to those who initially had such contact and originated the usage with that reference.

Thus, if a certain proposition is indeed singular, then there is a causal chain linking at least one of its occurring singular terms with some individual object. In speech or thought, the function of such logically proper names is paradigmatically (though not exclusively) performed by ordinary proper names. Yet because any of the latter may aim at picking out an individual object while failing to do so in fact (i.e., turning out vacuous), questions concerning whether tokens of declarative sentences containing putative logically proper names actually express singular propositions are a posteriori—requiring an investigation of the causal history of, e.g., occurring, ordinary proper names. Clearly, whether the referent of any singular term exists, with those who use it in their speech or thought having had causal commerce with that entity (or with other members of the community whose referential usage of the term is traceable back to that of those who have had such commerce) amounts to an empirical matter of fact, knowable only a posteriori.

Externalists need not be directly concerned with the controversy between Fregeans and new theorists about the correct semantic account of singular terms, and they may even follow traditional Fregeanism in taking ‘water,’ ‘tiger,’ and other natural- and biological-kind words to

be among the general terms whose logical function is that of predicates, with properties falling within their extension. Yet, as illustrated by standard twin-earth thought experiments, externalists must part company with the traditional semanticist in her claim that the extension of any such term could be entirely determined by its Fregean sense or mode of presentation. In this matter, their views are closer in some crucial respects to those of new theorists of reference. As often argued by the latter for the case of ordinary proper names, the externalist's thought experiments suggest that natural-kind terms also necessarily refer to the same thing in different possible worlds where they pick out anything at all—even when it is not an individual object but the essential property of some substance or species that falls under their extension instead. And of course, unlike logically proper names, natural-kind terms need not be considered purely referential, since Putnam (1975) has shown that they do seem to have meanings (though these fall short of Fregean senses). But if externalism is correct, then the metaphysical status of propositions containing natural-kind terms would be similar to that of singular propositions, in that the former seem likewise to “invoke” something in the environment. In their case, it is an essential property, external to the individual who entertains the proposition.

The analogy, however, can be extended to cover some epistemic features. Just as tokens of a singular term may fail to pick out a specific object, no doubt those of some predicate intended to pick out a certain external property could regularly fail to do so. Such failures of reference would occur in scenarios where there is no causal chain linking a putative natural-kind term with anything external to the speaker—i.e., when no essential property of a substance or species turns out actually to fall under the extension of tokens of that term. How is this possible? Recall that in the mid-1970s, not only did externalists reject senses as wholly determining the extension of natural-kind terms, but they also (e.g., Putnam 1975, Burge 1979) embraced the

direct theorist's causal account of reference, further developing it to become the full-fledged causal account we know today. That turned out to be no historical accident: without Fregean senses, some causal account of reference was needed if externalists were to avoid a magic theory of reference. In their account, for any putative natural-kind term to amount to a genuine term of that sort, there must be a causal chain linking the term conventionally available in the speaker's linguistic community with some actual substance or species in the environment. This requires that the speaker, in his direct commerce with a certain natural kind, has introduced the term to refer to that natural kind--or else that he has taken it up from others in his linguistic community whose referential use of it is traceable back to those who originated it in their causal transactions with a certain substance or species (cf. Evans 1973).

Yet, as argued above, given a plausible version of externalism, there is room for holding that terms intended to refer to some natural kinds may regularly fail to pick out the right sort of external property. A native of our world, when he sincerely asserts 'Ether is lighter than phlogiston,' expresses a proposition that picks out no property at all external to the speaker. No doubt intentional ascription in such cases is internalist, for any two replicas sharing identical properties from the skin inward (non-intentionally described) could have the thought *that ether is lighter than phlogiston* no matter how different their environments turned out to be. Since a plausible version of externalism can countenance cases of this sort, therefore, given that doctrine, to determine whether any propositional-attitude content is externally determined requires investigation of the causal history of its occurring terms, which is an empirical matter. For how could any such question be settled just by reasoning? I submit that if semantic externalism is correct, then the justification of belief about substantial entailments from words and thoughts to the environment rests (at least in part) on empirical investigation.

Furthermore, if one's faculty of self-understanding is properly functioning, then, given semantic externalism, it may appear on a certain epistemic view that one could come to know the implications of one's thought contents a priori. But since the notion of justification fueling the paradox of concern here is that of epistemic *internalism*, the temptation to hold that view is now completely removed (see note 5). If this is correct, there is no epistemic problem facing the attempt to hold externalism together with privileged self-knowledge, simply because the above triad may now be recast as follows:

- 1* Oscar can know a priori that his current propositional attitude involves a *putative* natural-kind term, *k*.
- 2* Oscar can know a priori that, if his current propositional attitude involves a putative natural-kind term *k*, then *k* may pick out a certain natural kind.
- 3 No one can know a priori that any natural kind exists.

Clearly, there is nothing inconsistent about the attempt to hold all of these claims at once. Yet it remains to be shown that the doctrine of privileged self-knowledge is preserved by (1)*, and that externalism does not sanction entailments stronger than (2) *. Let us begin with the latter, which requires a closer look at the semantic commitments of that doctrine.

IV

I have argued that standard twin-earth thought experiments suggest two things: first, that the referents of natural-kind terms partly determine their meaning, and, second, that in any possible world where tokens of those terms have reference at all, they necessarily pick out samples of exactly the same substances (or species) falling within their extension in the actual world. Given these theses, traditional Fregeanism cannot provide a correct account of natural-

kind terms. For on Fregean assumptions, the extension of any general term, a certain property, is entirely determined by the term's sense (or cluster of senses), cashed out as a *non-demonstrative* concept in the speaker's mind (that is, her *non-demonstrative* way of thinking about a certain property falling under the extension of the term). Yet this leads to the internalist conclusion that T-Oscar could have psychological attitudes involving the concept 'water' on Twin Earth, where there is no H₂O but only superficially identical XYZ, and where nobody has ever been in contact with speakers having such a concept. That would be the case if his non-demonstrative ways of thinking about the superficial qualities of XYZ were relevantly similar to Oscar's when he has words or thoughts involving the concept 'water,' even when it is not XYZ but H₂O what falls under the extension of Oscar's term on Earth. This internalist result follows inevitably from a semantics that takes the meaning of any predicate to supervene entirely upon local properties of the individuals who entertain them in their speech or thought (see, e.g., Searle 1983).

Needless to say, on semantic intuitions of this sort, some relatively common scenarios would become altogether mysterious: viz., those where speakers seem to succeed in having words and thoughts about certain natural kinds in spite of their profound misconceptions and errors concerning the designated substances.⁵ Externalists, on the other hand, have no problem in accounting for those scenarios. In fact, well known externalist thought experiments (such as Putnam's 'elm/beech tree,' 'water/t-water,' and 'aluminum/molybdenum' cases) were devised precisely to elicit the intuition that an individual whose conception of some natural kind is incomplete or partially mistaken may still be able to entertain, not just *de re* attitudes, but *de dicto* ones as well involving that natural kind (Putnam 1975; Burge 1982a, 1982b). What matters for externalist attribution of content is the existence of referential links between certain terms embedded in a psychological attitude and some substances and species in the environment. Links

of that sort are thought to rest, at least initially, on the causal commerce of speakers with paradigm samples of those substances and species. Once referential links have been established in that way, they may later be transmitted to others in the linguistic community through social interaction, with no further direct causal contact needed.

Yet further contact and expert knowledge could at any time produce changes in the speakers' ways of thinking about the extension of some such terms—as happened with 'fish,' when it came to be understood that porpoises and whales did not fall within its extension, or 'jade,' when gemologists discovered that there are in fact two distinct minerals, nephrite and jadeite, falling within that term's extension (Quine 1969, and Putnam 1975, respectively). Note that, if it is typically the causal contact of speakers with an essential property constitutive of the relevant kind that initiates a communal practice of using some natural-kind term with a certain extension, then the extension of any such term is fixed in a manner entirely independent of whether those who begin using it have an accurate conception of the property falling within its extension. Although experts in the community may be able to provide a scientific account of such properties, they are in no way required in order to secure the extension of natural-kind words. As often emphasized by causal theorists of reference (e.g., Kripke 1972; Putnam 1975), terms such as 'water' and 'gold' were used to pick out samples of substances with a certain chemical structure long before any expert was able to discern the chemical composition of those substances—just as some essential properties of whales and porpoises fell under the extension of tokens of 'whale' and 'porpoise' long before it was discovered that those species do not qualify as fish.

Since misconceptions and even ignorance about the essential properties of natural kinds and species, however widespread among speakers, seem not at all to undermine their success in

using conventionally available natural-kind terms to predicate those properties, this counts as evidence against Fregeanism, a theory standardly construed as lacking a causal account of reference. Externalists, on the other hand, offer precisely such an account, together with the view that the extension of genuine natural-kind terms necessarily remains the same in any possible world where terms of that sort pick out anything at all.⁶

V

Semantic theses of this sort suggest that the epistemic status of propositions containing genuine natural-kind terms must be altogether different from that of propositions involving only internally determined terms. For assuming such theses, how could either self-ascriptive beliefs about externally determined propositions, or belief in their externalistic entailments qualify for nonempirical justification of the type available, for example, when one believes *that triangles have three internal angles*, or *that bachelors are unmarried men*? Since the latter are prime candidates for a priori justification, if beliefs that ordinarily do not require investigation of the environment, even when they may presuppose some empirical propositions, are held to be a priori too, this must be under a different sense of that notion—as, for instance, when one thinks *that one has a headache*, or *that Jessie Jackson supported Al Gore in the 2000 US elections*. For each of these presupposes empirical propositions concerning, e.g., the existence of my head,⁷ and that of Jackson and Gore, respectively.

Such contrasting cases support a distinction between a strong notion of a priori justification (or knowledge) and a weak one, construed as follows:

a priori_s = A property of belief in propositions which, resting on no empirical assumption at all, are not open to challenge on a posteriori grounds.

a priori_w = A property of belief in propositions which, though *prima facie* justified without empirical investigation, rest on some empirical assumptions, and are thus open to challenge on a posteriori grounds.

Under normal circumstances, when I claim to know by introspection that I believe *that water is wet*, I take for granted that water exists, and my term ‘water’ refers to water—that is, in the absence of contrary evidence, I assume that my putative natural-kind term is genuine. But, given the semantic commitments of externalism, since beliefs involving such terms would ultimately rest on empirical propositions (even when ordinarily and in the absence of evidence to the contrary their justification does not require investigation of the environment), they all fall short of being a priori in the strong sense: i.e., they are *not* indefeasible on a posteriori considerations.

Yet if self-ascriptive beliefs about thought-contents are not available a priori in the strong sense, doesn’t that undermine their privileged epistemic status? Not at all, if that status stems, not from their apriority but from their being available to the first person with special access and authority—which, as we have seen, are both consistent with their being a priori only in the weak sense. Recall that, given special access, beliefs of that sort are directly justified—i.e., based on neither evidence nor inference. And, given first-person authority, there is a presumption that they are also predominantly true, and thus highly eligible to be counted as knowledge. Compare my belief that *I am thinking that water is wet* with my believing that--say--*water exists*. Arguably, both are in some sense empirical, but while the latter is neither truth-warranted nor direct, the

former has these properties and is therefore epistemically privileged (qualifying for a priori_w). This is why self-ascriptions of that sort are ordinarily trusted: noticing that they are in some sense epistemically special, we simply don't bother to investigate the environment before making knowledge and justification claims about them unless presented with convincing evidence to the contrary. In light of this, there is room for yet another recasting of the self-knowledge claim in the incompatibilist triad. Claim (1) may now be construed as holding that I can know a priori_w that my thought involves a certain term, 'water.'

Thus construed, the claim seems plausible. But our discussion appears to suggest that claim (2) might have a similar epistemic status, since externalism is knowable by philosophical argument. Given (1) and (2), then, I might still be in a position to know a priori (though only in a weak sense) that water exists. However, since the issue of whether or not the putative natural-kind term 'water' turns out to be genuine ultimately depends on an empirical question concerning the causal history of my tokens of that term, the specific externalist entailment in this case cannot run stronger than this,

2** If my thought involves the putative natural-kind term 'water,' and this term is a genuine natural-kind term, then there is (or has been) water in my environment.

It doesn't really matter that I may come to know a priori_w that my thought involves the term 'water,' since the antecedent of the externalist entailment in (2)** is in fact a compound proposition, with a conjunct unavailable a priori *in either sense* of that notion. For, clearly, whether my tokens of 'water' express a genuine natural-kind term depends on their causal history, something I could know with neither special access nor first-person authority. It follows that, given the semantic commitments of externalists, if entailments from propositional-attitude contents to environmental conditions need not be more demanding than (2)**, then (when

justified and true), they would amount to knowledge in some sense, but ultimately of the a posteriori type.⁸

VI

The incompatibilist's chief goal was to show that the attempt to hold externalism and privileged self-knowledge concurrently leads to an absurd conclusion. Suppose we grant her (as I think we should) that Oscar could figure out some general entailments from thought-contents to the world by knowing externalist theory (which he could come to know through standard twin-earth thought experiments) and the contents of his current thoughts.⁹ Given externalism and privileged self-knowledge, the consequent of such entailments appears to follow by simple deduction, without Oscar's conducting any investigation of the environment at all. Yet this objection cannot be made out if the semantic theory endorsed by externalists has the epistemic consequences suggested here. Since there is more than one notion of apriority at issue in this debate, externalists could insist that only beliefs presupposing no empirical proposition whatsoever can count as a priori in the more interesting sense of *nonempirical*. Even when self-ascriptive beliefs about propositional-attitude contents may be eligible for a priori justification, that would be only under a weaker construal of that notion. At the same time, belief about specific externalist entailments would qualify for neither.

Nothing absurd then can be deduced from such premises. For even after granting that one of the needed premises has the special epistemic status of being knowable without any special investigation of the environment, the conclusion required for the *reductio*--that externalism and privileged self-knowledge together entail that one could come to know a priori (in the sense of nonempirically) that certain natural kinds exist--does not follow. Such a conclusion would

require that the property of being knowable nonempirically be transmitted under known entailment, and this clearly cannot happen, since the premises themselves fall short of having any such epistemic property. Once externalism is properly construed, it becomes plain that the attempt to hold it concurrently with privileged self-knowledge generates no paradox at all.

Notes

1 . Goodman (1965) has argued that reflection which aims at reaching an accord between our rules of inference and the instances of inferences we accept can provide the only justification for the former. Stich (1991), however, is skeptical about what such a ‘virtuous’ circle could justify. For an assessment of the role of reflective equilibrium in justification, see Sosa 1991.

2 . Since here I assume that meaning and propositional-attitude content are analogous, ‘term’ should throughout this paper be taken to apply to either words or concepts, and ‘proposition,’ to the content of either a declarative sentence or a thought.

3 . Although Burge (1979) has argued that content externalism is compatible with Fregean semantics, he probably has in mind *neo*-Fregeanism of the sort proposed by, e.g., Wiggins (1993). Putnam (1975), on the other hand, famously urged that traditional Fregeanism entails “methodological solipsism,” which amounts to internalism about meaning and content.

4 . Note, however, that if ‘justification’ and ‘knowledge’ are construed as epistemic externalist notions, then no piecemeal empirical checking is needed, and this might be erroneously taken to suggest that self-knowledge claims are a priori in the strong sense of being nonempirical. A

reliabilist, for example, may hold that ascriptions of content to one's own propositional attitudes rest on the trustworthy faculty of self-understanding (as argued by Heil 1988, and perhaps by Burge 1998). If such a faculty is indeed reliable, then when one ascribes to oneself certain contents on the basis of understanding, one's claim could be justified and amount to knowledge without the need for piecemeal examination of the environment. To determine whether self-understanding is actually a reliable faculty, however, would itself require some empirical investigation involving, e.g., checking its track record. In any case, epistemic externalism need not be assumed here, since the types of epistemic notions that matter for the paradox of externalism and self-knowledge are internalist, requiring reflective justification (or, *knowing that one knows*). Suppose that in the "normal" case many propositional-attitude contents turn out to be externally determined, as some have pointed out (Brewer 2000, Davies 1998). Still, given epistemic internalism, the individual who entertains those contents may *not* know nonempirically that her thoughts involve contents of that sort, even when self-understanding could indeed be reliable.

5. As argued by direct theorists, not only was 'water' used to pick out H₂O long before the discovery of that property (when *a fortiori* speakers could have had *no* conception of it) but 'marsupial mouse' was once mistakenly thought to pick out a species kindred to mice (rather than to kangaroos and opossums, as we classify them today). And although a bat is not a flying mouse, it appears that German speakers once assumed that it was; hence their term 'Fledermaus.' While all these may prove puzzling to a theory that takes Fregean senses to determine the extension of such terms, they could be easily accommodated if the grounding of their extension were considered a matter of initial causal interaction with certain sources and social cooperation.

6 . At the same time, Burge (1982a) rejects the view that natural-kind terms might be similar to indexicals, as that would seem to lend plausibility to the narrow- vs. broad-content distinction that he wishes to eschew. For, given that analogy, natural-kind terms could be said to have a narrow meaning that like a Kaplanian character, may remain constant from world to world—while its broad meaning can vary from context to context with relevant changes in the physical environment. But that clearly falls short of the conclusion externalists standardly draw from their twin-earth cases. For objections to Burge’s reasoning here, see Sosa (1993).

7 . This example was suggested to me by Stephen Schiffer.

8 . McKinsey (1991a, 1994b) believes he has shown that taking externalist entailments to be metaphysically necessary yet knowable a posteriori would trivialize the characteristic thesis of externalism. But this is only because he construes that thesis in a way that is far too weak to capture what externalists have in mind. Thus construed, a blatantly trivial entailment such as that *Oscar's thinking that water is wet metaphysically entails that Oscar's biological parents existed* seems to satisfy that thesis. But when externalism is properly understood, it must be taken to make more specific dependence claims about content—e.g., holding that Oscar’s thought in that case necessarily depends in part upon the existence of a certain substance external to him, which is specifically related to the content of his thought. And there is no danger that this would trivialize the characteristic thesis of externalism, since even specific claims about the dependence of an individual’s thought contents upon certain environmental conditions may support general supervenience claims about the individuation of propositional-attitude content types, which are what externalists ultimately wish to make. See Brueckner 1995.

9. Given externalism, should propositions such as ‘If I have contentful thoughts, then something other than me must exist,’ be considered *prioris* or nonempirical? This complex question has little bearing on the paradox of concern here, which involves the possibility that externalism and privileged self-knowledge open a nonempirical way to know *more substantial* empirical propositions.