

EXTERNALISM AND THE RESOLUTION OF SELF-KNOWLEDGE

AMIR HOROWITZ* · HILLA JACOBSON**

I. INTRODUCTION

THIS paper suggests a new way for defending semantic externalism from what we take to be the most serious attack against it in the context of the discussion of the a priori nature of self-knowledge. Specifically, we shall argue that the resolution of our a priori knowledge of our beliefs on the assumption that their contents are externally determined is identical to the resolution that it makes sense to attribute to our knowledge of our beliefs independently of any assumption about content-determination.

Semantic externalism is the thesis that the contents of (at least some of) our mental states are determined, at least partly, by factors of our environment. These factors might be physical ones – the fact that the subject’s physical environment is a certain way (e.g., that the watery stuff on Earth is H_2O), or socio-linguistic ones (the fact that some people use some term to refer to some stuff), but at any rate, they are external to the mind of the individual whose contents are concerned. We shall focus on “physical” externalism, and specifically, on the externalism that is suggested by Putnam’s Twin Earth fantasy: the concept “water” in the thoughts of an Earthling refers to samples of H_2O , whereas the concept “water” in the thoughts of Oscar’s molecular and phenomenal duplicate who is an inhabitant of Twin Earth – a planet whose only differences from Earth are that its watery stuff is XYZ rather than H_2O and other differences that ensue from this difference – refers to samples of XYZ (See Putnam 1975).¹

The expression “self-knowledge” will be used here (as it is often used in discussions of the present issue) to refer to one’s knowledge of one’s own mental states, and in particular, to one’s knowledge of the contents of one’s own

* The Open University of Israel, 1 University Road, Ra’ Anana, Israel. E-mail: amirho@openu.ac.il

** Ben Gurion University of the Negev, Faculty of Humanities and Social Sciences, Department of Philosophy, P.O. Box 653, Be’er Sheva 84105 Israel. E-mail: hillaj@bgu.ac.il

¹ This paper of Putnam includes also an argument for socio-linguistic externalism. The other well-known argument for this latter genre of externalism is Burge’s (1979) arthritis argument.

mental states (to the knowledge, that is, of “mental contents”). Self-knowledge, thus understood, is supposed to be a priori, in the sense that it need not involve any empirical observation. In particular, self-knowledge is supposed to be achievable independently of any knowledge of the non-mental environment. It should be noted that not everyone who maintains that self-knowledge is a priori maintains that it is independent of any knowledge of the non-mental environment – one reaction to the alleged tension between externalism and the alleged a priori nature of self-knowledge consists in the claim that we can have a priori knowledge of the non-mental environment (see, e.g., Sawyer 1998). If the argument to be suggested in this paper is sound, then externalism provides no reason against the view that self-knowledge is achievable independently of any empirical knowledge *and* independently of knowledge of the non-mental environment, whether or not the latter two kinds of knowledge are in fact identical (though if we are right, externalism also does not provide any reason against their identification).²

The reason why there seems to be a tension between semantic externalism and the a priori nature that is attributed to self-knowledge is simple. If an environmental factor is a determinant of the content of a mental state, then knowledge of this content involves knowledge of this environmental factor, which knowledge is not a priori but depends on empirical observation (e.g., on research as to the microstructure of some stuff). Thus, the knowledge of an externally-determined content depends on empirical observation. Naturally, many take this result to be a problem for semantic externalism, even its *reductio*.³

It should be emphasized that the problem that is attributed to externalism according to this line of thought does not concern *the content determination* of one’s belief about one’s belief, but rather the epistemological status of this second-order belief. Suppose that Oscar believes that water quenches thirst (this is the belief that is standardly expressed in English by the sentence “Water quenches thirst”).⁴ According to externalism, this belief is (exclusively) about

² We are following the standard literature in presenting the problem in terms of a priori versus empirical knowledge, although we are not happy with characterizing introspection-based knowledge as a priori. In an important sense, such knowledge seems to us a sort of empirical knowledge: it is not based on reasoning, but rather, as a long philosophical tradition maintains, on an “inner perception”. But since nothing of essence in the present discussion depends on this point, we shall ignore it.

³ As McKinsey (1991) put it, externalism and the assumption that self-knowledge is a priori entail the absurd view that we have a priori knowledge of environmental. As noted, not everyone takes this view to be absurd. McKinsey himself rejects the assumption that we have a priori knowledge of our own, present mental states, and thus avoids the alleged absurd implication of externalism.

⁴ To avoid confusion, note that we are speaking here of how this belief is *expressed* and not of how it is *reported*. It would be reported by the believer by the sentence “I believe that water quenches thirst”, and this report also expresses the believer’s *second-order* belief.

samples of H_2O and their characteristics. Suppose, further, that Oscar also has the second-order belief that is standardly expressed in English by the sentence “I believe that water quenches thirst”. And suppose, still further, that Oscar has no knowledge of the chemical structure of water and cannot distinguish samples of H_2O from samples of XYZ. One might then wonder how, given Oscar’s ignorance, the concept “water” that figures in his second-order belief can be (exclusively) about samples of H_2O and thus enable this belief to be about a belief which concerns (exclusively) characteristics of samples of H_2O . This question about the content determination of second-order beliefs is *not* the question that lies at the heart of the issue of externalism and self-knowledge, and it is *not* the question with which we are concerned. The externalist has no special difficulty to treat this semantic question, as Burge (1988) shows. The idea is that according to semantic externalism the contents of our second-order beliefs are determined by external factors in exactly the same way that the contents of our first-order beliefs are determined. The very same fact that makes the concept “water” that figures in that first-order belief of Oscar the concept of samples of H_2O and makes that belief itself be concerned (exclusively) with samples of H_2O – namely the fact that the watery stuff in Oscar’s environment is H_2O , also makes the concept “water” that figures in that second-order belief of Oscar a concept of samples of H_2O and thus makes that second-order belief a belief about a belief about such samples, Oscar’s ignorance notwithstanding. According to semantic externalism, one need not know anything about the environment (in particular, about samples of H_2O) in order to have a (second-order) belief about a belief about samples of H_2O exactly as one need not know anything about the environment (in particular, about samples of H_2O) in order to have a (first-order) belief about samples of H_2O . It is the essence of the externalist idea that factors in one’s environment that are not known to one participate in determining the contents of one’s beliefs. One only has *to be* in the right environment (and to be appropriately related to samples of the relevant stuff) for having a certain belief. And since the environmental condition for having a certain first-order belief is also the environmental condition for having a belief about that belief, one need not gaze outside in order to have that second-order belief. Thus, semantic externalists can account for the contents of second-order beliefs. One can form beliefs about one’s (first-order) beliefs whose contents are externally determined without gazing outside.

2. SELF-KNOWLEDGE AND DISTINGUISHABILITY

Still, it is argued that such second-order beliefs, when arrive a priori, cannot constitute knowledge. What underlies this view? The idea is that Oscar’s belief that is expressed by “I believe that water quenches thirst” – a belief that is

exclusively about samples of H_2O – cannot constitute knowledge if arrived at a priori, because Oscar *cannot distinguish* a priori between a situation in which his first-order belief concerns H_2O and a situation in which it concerns XYZ. This claim is based on the epistemological view that knowledge requires distinguishability: that one cannot know that P if one cannot distinguish situations that instantiate P from those that do not. A person who believes (truly) that she sees a German Shepherd but cannot tell a German Shepherd from a Siberian Husky does not know that she sees a German Shepherd. Oscar seems to be in exactly such a situation vis a vis his first-order beliefs, hence he cannot know their contents (see, e.g., Boghossian 1989). This is what Brown (2004) calls “the discrimination argument”.⁵

Does knowledge requires distinguishability of that sort? Specifically, does one have to be able to distinguish the object of one’s belief from any other object for one’s belief to constitute knowledge? Some philosophers reject the claim that Oscar cannot know the content of his first-order belief, on the grounds that the epistemological view in question is false. If the reliabilist approach to knowledge is correct, that second-order belief of Oscar does (if true) constitute knowledge, for it results from a reliable process, that is, from a process (namely introspection) that tends to produce true beliefs. Some other philosophers argue that one’s knowledge of a proposition requires that one be able to distinguish it only from some relevant alternatives, and that this requirement can be met in the case of one’s beliefs about beliefs whose contents are externally-determined even if one is confined to a priori resources.⁶

The parties to this controversy accept the idea that the answer to the question whether mental contents that are externally determined can be known a priori depends on which theory of knowledge is correct. This in itself is plausible; in some sense it is even trivial. But in the framework in which answers to the question whether mental contents that are externally determined can be known a priori are taken to bear on the plausibility of externalism, relying on such a linkage between the a priori knowability of such beliefs and theories of knowledge seems to us problematic. For it seems implausible that the plausibility of externalism depends on which epistemological theory is correct. The crucial question as far as the plausibility of semantic externalism in light of epistemological considerations is concerned is not whether, assuming that semantic externalism is correct, Oscar’s second-order belief may be justly called “knowledge”. The crucial question is whether what externalism implies about such second-order beliefs is indeed the case, regardless of whether or

⁵ Brown distinguishes this argument from “the illusion argument.” We believe that our treatment of the discrimination argument is applicable also to the illusion argument, but discussing the illusion argument is beyond the scope of this paper.

⁶ See, e.g., Falvey and Owens (1994), McLaughlin and Tye (1998), Brown 2004.

not such beliefs may be said to constitute knowledge. One might think that externalism would be in trouble if – given the correct theory of knowledge – it implies that we cannot know the contents of our beliefs a priori, since we do know them a priori. But that in fact means that what the opponent of externalism has to show is that we know the contents of our beliefs *in a sense in which* externalism implies that we don't; and that means that what matters is whether the actual epistemic relations between our first-order beliefs and our second-order beliefs are or are not those that externalism implies, regardless of the title and of the epistemic merit that we attribute to the relevant second-order beliefs.

Since knowledge in the sense whose absence is said to undermine externalism according to the objection in question involves distinguishability, the charge that externalism faces a difficulty in this regard should consist of two claims: the claim that Oscar cannot distinguish a priori between the content that his (first-order) belief has according to externalism and some other contents (e.g., such that concerns samples of XYZ), and the claim that Oscar thus fails to meet the epistemic standard that we have reason to assume – regardless of the externalism/internalism issue – people do meet. If these two claims are correct, externalism is in trouble for implying something that violates the epistemology of the mental. Herein, we believe, lies the real epistemological difficulty that externalism faces.

3. SELF-KNOWLEDGE AND THE QUOTATION PRINCIPLE

Are those two claims correct? We grant the first claim. Let us focus on the second claim, the claim that in failing to distinguish a priori between the content that his (first-order) belief has according to externalism and some other contents (e.g., such that concerns samples of XYZ), Oscar misses something that we have reason to assume – regardless of the externalism/internalism issue – that people do not miss. We shall now argue that this claim is false: the “resolution” of one's (a priori) responsiveness to one's first-order beliefs on the assumption that their contents are externally determined is *identical* to the resolution that it makes sense to attribute to one's (a priori) responsiveness to one's first-order beliefs independently of any assumption about content-determination. If this is correct, then externalism meets the challenge that is presented to it by the discrimination argument.

Let's see what this means. Suppose that Noah believes that it's raining – that he believes that belief that he would express by saying “It's raining.” If no cognitive malfunction occurs, his relevant second-order belief is the belief that he believes that it's *raining* – a belief he would express by saying “I believe that it's raining.” He would not believe that he believes that it's *snowing*: he would not have (in virtue of his rain belief) a second-order belief he would express by

saying “I believe that it is snowing.” We may say that Noah’s responsiveness to the contents of his first-order beliefs is such, that these contents are described in his verbal expressions of his first-order beliefs and of his relevant second-order beliefs by means of the same words (except for the “I believe that” operator). If his first-order belief is the one he would verbally express by saying “P”, then his relevant second-order belief is the one he would verbally express by saying “I believe that P”. We may call this principle that Noah’s second-order beliefs obey “the quotation principle”. It is important to note that ascribing the quotation principle to the relations between those first-order and second-order beliefs of Noah does not depend on whether or not the contents of his first-order beliefs are externally-determined. It seems to be theory-neutral.⁷

It might be that the quotation principle implicitly underlies the view that the “resolution” of one’s responsiveness to the contents of one’s first-order beliefs is “maximal”, that is, that one can distinguish the contents of one’s beliefs from all other possible contents: for according to this principle, if one believes the belief one would express by “P”, then if no malfunction occurs, one’s second-order belief wouldn’t be the belief one would express by “I believe that Q”, nor would it be the belief that one would express by “I believe that R”, and so on. One would be able to distinguish one’s first-order beliefs, it seems, from all other possible first-order beliefs.

But unless externalism is ruled out, it is not legitimate to move from the quotation principle to the view that the resolution of people’s responsiveness to their first-order beliefs is maximal. Consider Oscar’s belief that water quenches thirst, a belief he would express by saying “Water quenches thirst.” His relevant second-order belief is the belief that he believes that water quenches thirst, a belief he would express by saying “I believe that water quenches thirst.” Now if semantic externalism is true, and specifically, if the content of “water” is externally determined in the way suggested by the Twin Earth argument, then as long as Oscar is confined to a priori resources, the resolution of his responsiveness to his first-order beliefs is not maximal. We already saw this: although his second-order belief is about a (first-order) belief that is exclusively about samples of H₂O, he cannot distinguish – his second-order

⁷ The quotation principle is put in terms of the *verbal expression of first-order and second-order beliefs*, but this is not to say that it is not a substantial principle. The appeal to the way those beliefs are expressed is a convenient way to refer to the nature of those beliefs themselves: according to the principle, the second-order beliefs one normally has concerning one’s first-order beliefs are those second-order beliefs whose verbal expressions relate to the *verbal expressions of one’s first-order beliefs* in a specific way (the way the principle specifies). Thus, the quotation principle in fact controls the relations between first-order beliefs and second-order beliefs and not only the relations between their respective verbal expressions. Truly, there is an essential linguistic dimension to the quotation principle, since it can only pertain to mental contents that are expressible in language – we deal with this issue below.

belief does not embody a distinction – between this first-order belief and a belief that is exclusively about samples of XYZ. However, Oscar’s second-order belief obeys the quotation principle: this belief and the first-order belief it is about are expressed by the same words – “water quenches thirst” (except for the “I believe that” operator). So unless externalism is ruled out, accepting the quotation principle as the principle that dictates what second-order beliefs one would normally have, given one’s first-order beliefs (accepting, that is, that the quotation principle controls the relations between first-order beliefs and second-order beliefs about them when no malfunction occurs) does not require us to accept the view that people’s responsiveness to their first-order beliefs is maximal; it does not require us to accept that one can distinguish one’s first-order belief from any other possible belief. The maximal resolution view is stronger than the quotation principle. Since this is true unless externalism is ruled out, then also in cases such as that of Noah, whose beliefs’ contents are not assumed to be internally determined, we cannot conclude from the fact that Noah’s above-mentioned second-order belief obeys the quotation principle that the resolution of this belief’s responsiveness to the relevant first-order belief is maximal.⁸

Further, it is not only that the quotation principle does not entail the view that the responsiveness of second-order beliefs to first-order beliefs enjoys maximal resolution. Rather, we have a reason to reject this view. We have a reason to reject the maximal resolution view of the relations between first-order and second-order beliefs since it is not only that the quotation principle is true of these relations. It is the principle that controls those relations (when no malfunction occurs), or at any rate, *no stronger principle control those relations*. That is, differences among second-order beliefs cannot be more sensitive to differences among first-order beliefs than they are supposed to be according to the quotation principle. The principle, recall, states that if one’s first-order belief is the belief one would verbally express by saying “P”, then one’s relevant second-order belief is the belief one would verbally express by saying “I believe that P”. The principle is formulated in terms of the verbal expressions of the beliefs in question. Now as far as the beliefs’ verbal expressions are concerned, this principle states the strongest possible sensitivity of people to their (first-order) beliefs. For since on this principle the second-order belief is such that its verbal expression employs the same phrase that is employed by

⁸ We do not and need not argue that *there are exceptions to the maximal resolution view that are not concerned with externally-determined contents*. Externalism faces a problem if the epistemology of externally determined contents is found to be different from that of contents that are not assumed to be externally determined, but not if the epistemology of externally determined contents is only found to be different from that of internally determined contents. See below.

the verbal expression of the first-order belief (it employs this phrase plus the belief operator), then the second-order belief is responsive to every content feature of the first-order *belief that is expressible in language*.⁹ This being so, the only hope for one who wishes to avoid the claim that the quotation principle states the strongest possible responsiveness of people to all content features of their (first-order) beliefs is to adhere to the view that the contents of our beliefs include elements that are ineffable.

Before addressing this option, it will help to make our argumentation explicit. We argue that:

1) No principle stronger than the quotation principle controls the relations between first-order beliefs and second-order beliefs about them.

2) In cases of thoughts whose contents are externally determined, the relations between first-order beliefs and second-order beliefs about them conform to the quotation principle, even if the second-order beliefs in question are arrived at a priori.

3) (From (1) and (2)) Externalism is not committed to taking people to be less sensitive to their first-order beliefs than they in fact are, and hence it does not face a problem in this regard.¹⁰

The idea, in short, is that the sensitivity of self-knowledge is described by the quotation principle, and since this principle is also valid in cases of externally-determined mental contents, externalism does not imply that we are less sensitive to our mental contents than we in fact are. We saw that premise (2) is straightforward. As to premise (1), we argued that it is true as far as the verbal expression of content is concerned. In other words, if the contents of beliefs are verbally expressible, then no principle stronger than the quotation principle controls the relations between first-order beliefs and second-order beliefs about them. So if the claim that the contents of beliefs are (completely) effable is warranted, (3) is warranted. And note that this claim is concerned with the effability of *the contents* of beliefs; not that of the bearers of the contents, not that of their phenomenality (if they have phenomenality); and note, further, that the contents in question are *conceptual* contents. Bearing

⁹ One cannot object to this idea by claiming that the verbal expression of the concept “water” should be expressed in more detail than it is expressed by the word “water”, that is, by some detailed description of water. For however detailed this verbal description is, it should figure in the expressions of both the first-order belief and the second-order belief in which this concept figures.

¹⁰ In a way, our argument can be described as an argument to the effect that Burge’s idea of the content-determination of second-order beliefs applies to the epistemological dimension: that the Burgean resolution of second-order beliefs whose contents are externally determined suffices to defend externalism from the discrimination argument. Burge’s treatment of this epistemological dimension (1988) is along different lines. We think his treatment is mistaken, but cannot discuss it here (see, e.g., Brown’s (2004) critique of Burge).

these points in mind, this claim seems highly plausible. And then, we have firm grounds to accept the conclusion (3).

But let us not rule out the possibility that there are beliefs whose contents are ineffable.¹¹ Our opponents must appeal to contents that satisfy a strange conjunction of conditions: being ineffable, and at the same time being the objects of (self) knowledge, or at any rate of (second-order) beliefs. If there are ineffable contents, it seems to us that the only model on which they can be the objects of second-order beliefs is this: some element (and not merely concepts that apply to certain elements) figures in the content of a first-order belief, and this same element also figures in the content of a second-order belief about that first-order belief. However, externalists have no problem accommodating this model – it seems tailor-made for externalism. If *an external object* figures in a first-order belief, it also figures in the second-order belief about this belief.¹² So externalism does not suffer from any epistemological disadvantage regarding ineffable contents. And we are still entitled to the view that externalism is not committed to taking people to be less sensitive to their first-order beliefs than they in fact are, and hence it does not face a problem in this regard.

It is important to make one point more explicit. We do not deny that the resolution of one's (a priori) responsiveness to one's first-order beliefs is maximal in all but (supposed) cases of externally determined contents. That may well be the case, but it does not tell against externalism. In order to undermine externalism on the epistemological grounds under discussion, it has to be shown that the resolution in question is maximal regardless of whether content determination is internalistic or externalistic. There is no reason to suppose that the epistemological principles that apply to internally-determined contents apply to contents regardless of how they are determined. Now one might reject this claim and argue that the resolution of one's responsiveness to one's internally-determined contents is maximal, and that it is maximal for a reason that applies to contents regardless of how they are determined: it is simply that we must be able to distinguish a priori between any two contents *that differ in intension*, since the intensions of our beliefs are epistemically given – or, we may say, transparent – to us. However, this reason is not independent of the internalism/externalism issue. Whatever exactly intensions are, for the preset purpose they must be understood to be (also) determiners of

¹¹ Fodor (1987) takes *narrow content* (that is non-referential) to be ineffable, but what matters for the present issue is the effability of referential contents.

¹² Recall the Burgean point about the individuation of externally-determined second-order beliefs.

It is not that we believe that such (object-dependent) beliefs are ineffable in any significant sense, but our point does not depend on this view.

extensions, and externalism denies that intensions thus understood are “in the head” and epistemically given to us.

Perhaps this point reveals the source of the conviction that supposedly externally-determined mental contents violate some general epistemological principle: our traditional ideas of self-knowledge of mental contents – as, e.g., “maximal” in the sense in question – are based on a semantically internalist picture (and what enables us to hold on to these ideas is the fact that they are not tested against cases of externally-determined contents). Once internalist semantics is abandoned or put into question, it might escape our notice that the epistemology that is based on it is indeed based on it, and so this epistemology might still be taken for granted; but of course, it should not be taken for granted if semantic internalism is not taken for granted, and so it should not be used as evidence against semantic externalism.

ABSTRACT: This paper suggests a new way for defending semantic externalism from what we take to be the most serious attack against it in the context of the discussion of the a priori nature of self-knowledge. We shall argue that the resolution of our a priori knowledge of our beliefs on the assumption that their contents are externally determined is identical to the resolution that it makes sense to attribute to our knowledge of our beliefs independently of any assumption about content-determination. We shall also suggest what might be the source of the conviction that supposedly externally-determined beliefs violate some general epistemological principle.

KEYWORDS: epistemology, externalism, self-knowledge.

REFERENCES

- BOGHOSSIAN, P. 1989. “Content and Self-Knowledge”. *Philosophical Topics* 17: 5-26.
- BROWN, J. 2004. *Anti-Individualism and Knowledge*. Cambridge (Mass.): MIT Press.
- BURGE, T. 1979. “Individualism and the Mental”. *Midwest Studies in Philosophy* 4: 73-121.
- BURGE, T. 1988. “Individualism and Self-Knowledge”. *Journal of Philosophy* 85: 649-63.
- FALVEY, K. and OWENS, J. 1994. “Externalism Self-Knowledge, and Skepticism”. *Philosophical Review* 103: 107-37.
- FODOR, J. 1987. *Psychosemantics*. Cambridge (Mass.): MIT Press.
- MCKINSEY, M. 1991. “Anti-Individualism and Privileged Access”. *Analysis* 31: 9-16.
- MCLAUGHLIN, B. and TYE, M. 1998. “Content Externalism and Privileged Access”. *Philosophical Review* 107: 349-80.
- PUTNAM, H. 1975. “The Meaning of ‘Meaning’”. In his *Mind, Language and Reality: Philosophical Papers Vol. II*. Cambridge: Cambridge University Press.
- SAWYER, S. 1998. “Privileged Access to the World”. *Australasian Journal of Philosophy* 76: 523-33.