ISSN 1808-1711 doi: 10.5007/1808-1711.2015v19n2p197

WHY PRIVILEGED SELF-KNOWLEDGE AND CONTENT EXTERNALISM ARE COMPATIBLE

SERGIO ARMANDO GALLEGOS

Abstract. In the last twenty-five years, several authors have raised problems to the thesis that privileged self-knowledge is compatible with content externalism. In particular, the 'slow-switching' argument, which was originally put forth by Paul Boghossian (1989), aims to show that there is no satisfactory account of how we can have privileged knowledge about our own thoughts given content externalism. Though many philosophers have found ways to block the argument, no one has worried to address a major worry that Boghossian had when he presented the argument, which is to understand under which conditions privileged self-knowledge is possible given content externalism. In this paper, I offer a diagnosis of why the 'slow-switching' argument fails and I show how the diagnosis enables us to provide a partial response to Boghossian's worry.

Keywords: Privileged self-knowledge; content externalism; compatibilism; slow-switching.

1. Introduction

In recent decades, many epistemologists have accepted content externalism, which is characterized by the thesis that the contents of some of our thoughts are partially determined by the relations they bear to our external environment. Though there are distinct versions of content externalism that differ from one other as they stress different aspects of the dependence of our thoughts on various features of our environment,[1] all versions agree on the fact that the contents of our thoughts are not exclusively fixed by intrinsic features such as the microphysical structure of an individual's brain or other characteristics that are internal to him (i.e., they are not individuated individualistically). Now, one remarkable trait of many content externalists (e.g. Davidson 1987; Burge 1988 and Heil 1988) is that they also endorse the traditional view that we have privileged self-knowledge, which involves the thesis that we typically know better what we think when we are thinking it than any other people do (and conversely) because we have a privileged access to our own thoughts that does not depend on any empirical observations of our behavior.[2]

This simultaneous endorsement of content externalism and the privileged self-knowledge thesis is *prima facie* not surprising since both positions are well supported by compelling reasons. For instance, content externalism is strongly motivated both by considerations that its most vigorous detractors recognize[3] and by many well-known arguments based on thought experiments.[4] Moreover, the privileged self-knowledge thesis is also motivated by several observations that even its hardline

Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

critics admit (for instance, by the observation that, in a lecture, a speaker typically has better access to what he says and is seldom surprised by it, unlike his listeners, who may be frequently surprised)[5] and by well-known arguments such as those presented by Descartes in the *Meditations*. However, some philosophers have argued that serious difficulties arise if one accepts both theses. For instance, McKinsey (1991) has claimed that accepting both theses leads to absurd consequences (e.g., accepting that one can have privileged knowledge of facts that can at first sight only be known empirically). And Boghossian (1989) has contended that, if we accept content externalism, it is impossible to explain how we can have privileged self-knowledge.

These are quite formidable challenges. However, despite the problems that McKinsey and Boghossian have articulated, I believe that we should resist giving up either thesis since there is much at stake. For instance, we should not give up the thesis that we have privileged self-knowledge because, as Boghossian (1989, p.6) himself recognizes, 'self-knowledge is not an optional component of our self conception [...] It is a fundamental part of that conception, presupposed by some of the concepts (consider intentional action).' In a similar way, we should not give up content externalism because doing so puts us in an awkward position where we cannot explain how we can refer to objects in our environment given that, as Putnam (1981, p.16–7) contends, '[...] one cannot refer to certain kinds of things, e.g., *trees*, if one has no causal interaction with them or with things in terms of which they can described'.[6] In virtue of this, it would be desirable to develop a comprehensive theory of mental content that integrates fully both privileged self-knowledge and content externalism.

My project in this paper is not *this*, but a more modest one that can be used nonetheless as a stepping stone towards that goal. I present here a response to Boghossian's 'slow-switching argument', which presents a paradox to advocates of content externalism who also endorse the privileged self-knowledge thesis. Now, even though several responses to Boghossian's argument have appeared in the last twenty-five years (in particular, Falvey & Owens 1994; Vahid 2003 and Morvarid 2012), they all exhibit a common shortcoming: although they block the argument, they are all silent on an issue that worried Boghossian (1989, p.6) when he first presented the argument, which is the demand to gain *understanding*: 'I hope that by getting clear on the conditions under which self-knowledge is not possible, we shall better understand the conditions under which it is.' Thus, my goal here is not only to offer a response to the slow-switching argument but also to explain (at least partially) under which conditions privileged self-knowledge is possible given content externalism, which is a task that Boghossian confessed eluded him at the time he developed the argument.[7]

I proceed in the following manner. In section 2, I rehearse the assumptions that Boghossian makes about the nature of self-knowledge. Having done that, I present in section 3 the slow-switching argument as Boghossian formulates it. In section 4,

I offer a critical analysis of the argument and suggest a diagnosis of what is wrong with it. Subsequently, in section 5, I consider other responses to the argument and I show how my response, which shares some similarities with the diagnosis that Bernecker (2004) presents of another argument offered by Boghossian (the 'memory argument'), is in a sense better. In section 6, I consider some possible objections to the proposal that I make here and I respond to them. Finally, I provide a brief conclusion in section 7 that outlines some future lines of work.

2. Some Preliminary Remarks Regarding Self-Knowledge

Before presenting the argument, it is important to have a clear picture of the specific notion of privileged self-knowledge that Boghossian has in mind, since this will be crucial to analyze his argument. Boghossian initially remarks that the notion of privileged self-knowledge that he wants to defend is a notion of *knowledge*. Considering this, he (1989, p.6) writes that "by 'self-knowledge', I shall mean not just a true belief about one's thoughts, but a *justified* one." This shows that Boghossian accepts that, in order to have knowledge of one's own thoughts, at least three minimal conditions must be fulfilled: one must have beliefs about one's thoughts, these beliefs must be true and they must be justified.[8]

In addition to this observation, Boghossian makes a second crucial remark about the notion of self-knowledge. Since knowledge is typically considered to be a type of propositional attitude, several philosophers believe that one must think of instances of self-knowledge as a type of mental representations that belong to a 'language of thought' in which mental representations are complex symbols that have certain syntactic properties that correspond to their semantic contents. In contrast to them, Boghossian (1989, p.6) questions this assumption since 'a language of thought model implies that there are *type-type* correlations between certain purely formal and intrinsic properties of thoughts and their semantic properties.' Since he recognizes that 'this is a heady assumption that stands to affect profoundly the account we are able to give of our capacity to know the semantic properties of thoughts', Boghossian rejects it.

Now, though the notion of self-knowledge that he accepts is robust enough to involve at least the presence of justified true beliefs in instances of privileged self-knowledge, Boghossian (1989, p.6) observes that numerous philosophers throughout history have endorsed notions of self-knowledge that are far too strong insofar as they rest on 'extravagant claims [...] made about our capacity to know our minds.' For instance, he points out that Descartes maintained that our self-knowledge was both infallible and exhaustive. In order to steer away from such immoderate positions, Boghossian introduces several further constraints that a satisfactory notion of self-knowledge must satisfy. The first two constraints are presented in the following passage:

> [...] self-knowledge is fallible and incomplete. In both the domain of the mental as that of the physical, events may occur of which one remains ignorant; and, in both domains, even when one becomes aware of an event's existence, one may yet misconstrue its character. (1989, p.19)

As we can appreciate, Boghossian makes clear here that our self-knowledge is, *pace* Descartes, compatible both with the possibility of *error* and with that of *ignorance*: we have privileged knowledge of our own thoughts even if we may be mistaken or ignorant about them. Moreover, Boghossian also remarks that having self-knowledge is not only consistent with the possibilities of ignorance and error, but also with the possibility of *doubt* created by global skeptical possibilities:

> The ordinary concept of knowledge appears to call for no more than the exclusion of 'relevant' alternative hypotheses (however exactly that is to be understood) and mere logical possibility does not confer such relevance. Similar remarks apply to the case of self-knowledge. (1989, p.12)

In addition, Boghossian introduces two other constraints that a satisfactory notion of self-knowledge must satisfy. Since he remarks that, with respect to one's own attributions of self-knowledge, 'I do not defend my self attributions; nor does it *normally* make it sense to ask me to do so' (1989, p.7, my emphasis), having self-knowledge is compatible with the possibility of correction from a third-person perspective if the circumstances are abnormal (e.g., if the individual in question happens to be deceived by an evil scientist). Finally, the last constraint that, according to Boghossian, a satisfactory notion of self-knowledge must fulfill is expressed in the following passage:

> Knowledge that is not a cognitive achievement would be expected to exhibit certain characteristics — characteristics that are notably absent from self-knowledge. For instance, [...] you would not expect cognitively insubstantial knowledge to be subject to direction: how much you know your own thoughts should not depend on how much attention you are paying to them. And yet, it does seem that, within bounds, self-knowledge can be directed [...] (1989, p.19)

What this passage shows is that self-knowledge must be taken as a cognitive achievement: knowing our own thoughts is often something that does not come for free, but a state that requires effort and attention from our part.[9] In light of all this, we can then see that the notion of self-knowledge that Boghossian has in mind possesses the following characteristics:

(a) It requires having (at least) justified true beliefs about our own thoughts.

(b) It is independent from a language of thought hypothesis.

(c) It is compatible with the possibility of error.

(d) It is compatible with the possibility of ignorance.

(e) It is compatible with the possibility of global skeptical doubt.

(f) It is compatible with the possibility of correction.

(g) It is a cognitive achievement.

These constraints are of crucial importance because, as I show below, they block certain responses that can be given to the main question that Boghossian is concerned with — namely, determining under which circumstances privileged self-knowledge is possible given content externalism. Having clarified which constraints a notion of privileged self-knowledge must fulfill for Boghossian, I turn now to a reconstruction of his argument.

3. A Tentative Reconstruction of the Slow-Switching Argument

The slow-switching argument that Boghossian presents is inspired by other arguments previously developed by Putnam (1975) and Burge (1988). In the case of Putnam, his argument, which aims to establish that the semantic contents of some of our thoughts (in particular, those involving natural kind concepts such as *water*) are partially determined by certain environmental features, involves the following thought experiment. Putnam asks us to imagine *two* individuals (Oscar and Toscar) who are duplicates in all microphysical, functional and psychological respects. The only difference between them is that, while Oscar lives on Earth where he has only contact with water, Toscar lives in another planet (called Twin Earth) that differs from Earth solely in the fact that it contains, instead of water, a liquid superficially identical with water but with a different chemical composition (let us call this substance 'twater'). In addition, Putnam assumes that Oscar and Toscar are laymen incapable of distinguishing water from twater at the chemical level.

After considering these two individuals living in two different environments, Putnam asks us to suppose that each individual travels to the other's respective planet. On the basis of this, Putnam then argues that it is plausible to maintain that, if the two individuals were to utter shortly after their travels 'That is water' while pointing to the liquid flowing from an open faucet in their respective new environments, both would be making mistakes. Indeed, when Oscar makes the utterance, Putnam maintains that he is presumably in a certain psychological state that involves the concept WATER, but this state fails to correspond to the extension of the term 'water' on Twin Earth, and a similar problem arises for Toscar. In virtue of this, Putnam (1975, p.144) concludes that, since the extensions of natural kind terms such as 'water' (which he identifies with their meanings) do not depend on our psychological states, "'meanings' just ain't in the head."

Using Putnam's thought experiment, Burge develops a similar argument. However, instead of supposing that two individuals travel from Earth to Twin Earth and *vice versa*, Burge considers the possibility that only *one* individual is shuttled back and forth between both planets in accordance to the following conditions:

> Suppose that one underwent a series of switches between actual earth and twin earth so that one remained in each situation long enough to acquire concepts and perceptions appropriate to that situation. Suppose there are occasions where one is definitely thinking one thought, and other occasions where one is definitely thinking its twin. Suppose also that the switches are carried out so that one is not aware that the switch is occurring. (1988, p.652)

Given these suppositions, Burge (1988, p.653) acknowledges that the upshot of the thought experiment is that 'the person would have different thoughts under the switches, but the person would not be able to compare the situations and note when and where the differences occurred.' Considering this, Burge then observes that some may construe this as suggesting that the person in question could not possibly know his own thought without making an empirical inquiry — a claim that he deems absurd.[10]

An important point to bear in mind here is that, when Burge introduces the thought experiment, he considers it as a mere logical possibility rather than as a relevant alternative. Though an individual could certainly undergo a series of switches between Earth and Twin Earth meeting all the conditions that Burge introduces, Burge (1988, p.654–5) makes clear that the slow-switching scenario is just a logical possibility that does not undermine the individual's privileged self-knowledge because 'in saying that a person knows, by looking, that there is [water] there [...] we also do not require that the person be able to recognize the difference between [water] and every other imaginable counterfeit that could have been substituted.'

Although the slow-switching scenario that Boghossian presents to motivate his argument is deeply influenced by Burge's thought experiment, Boghossian introduces an important change: the slow-switching scenario is taken to be not a mere *logical possibility* but rather a *relevant alternative* by making the following assumption:

> Imagine that Twin Earth *actually exists* [...] I invite you to consider then a thinker S who, quite unaware, has been shuttled back and forth between Earth and Twin Earth, staying long enough to acquire the concepts appropriate to his situation, and at the expense of the concepts appropriate to his previous situation. (1989, p.13, my emphasis)

Once the slow-switching scenario is turned into a relevant alternative by supposing that Twin Earth is an actual planet, Boghossian then proceeds to introduce the slow-switching argument in the following passage:

> Does S knows what he thinks while he is thinking it? Suppose he is on twin-earth and thinks a thought that he would express with the words 'I [*want water*]'. Could he know what he thought? The point to bear in mind is that the thought I [want water] is now a relevant alternative. He, of course, is not aware of that, but that does not change matters. Epistemic relevance is not a subjective concept. [...] S has to be able to exclude the possibility that his thought involved the concept [*water*] rather than the concept [*twater*] before he can be said to know what his thought is. But this means that he has to reason his way to a conclusion about his thought; and reason to it, moreover, from evidence about his external environment which, by assumption, he does not possess. How, then, can he know his thought at all? (1989, p.13–4)

Let me now offer a reconstruction of the argument. In light of Boghossian's initial remarks about privileged self-knowledge, it is clear that the first sentence expresses a rhetorical question since Boghossian takes the claim that we have privileged self-knowledge to be a non-negotiable assumption. Subsequently, he supposes, using his own version of the Twin Earth thought experiment, that an individual S is shuttled back and forth between Earth and Twin Earth in such a way that certain conditions are met. Having done this, he asks to us suppose that S remains in a certain place where he thinks a certain thought. At this stage, he contends that, if S knows his thought (which he does *ex hypothesi*), he should be able to rule out the relevant alternative that he is thinking the twin thought. But, since this requires doing empirical investigation, it then follows that S does not have privileged self-knowledge:

(A) S knows his thoughts in a privileged way. (Assumption)

(B) Suppose S is transported back and forth between Earth and Twin Earth (which is an actual planet) in such a way that

  (i) he is unaware of the shifts,

  (ii) his qualitative life and internal structure remains the same and

  (iii) he remains long enough in a place before the next shift to acquire the relevant concepts at the expense of the twin concepts.

(C) Suppose that, after constant switching, S remains on Twin Earth where he happens to think a twater-thought.

(D) If S knows his thought (which he does *ex hypothesi*), he must be able to rule out the relevant alternative that he is thinking a water-thought.

(E) In order to rule out the relevant alternative, S must investigate his environment.

(F) But, if S has to investigate his environment to know his thought, he does not have privileged knowledge of it.

(G) Thus, given (B) through (F), S does not know his thought in a privileged way.

On this basis, I offer in the next section a critical analysis of the argument. However, before doing that, I want to stress again a crucial point: Boghossian (1989, p.6) does not intend the argument to be a *reductio* of (A) since 'there can be no question of accepting the skeptical claim [about self-knowledge]'.[11]

4. Examination of the Argument: a Diagnosis

In order to offer a critical analysis of the argument, let us consider the initial assumption that S has privileged knowledge of his own thoughts. If S has indeed privileged knowledge of his own thoughts, he must then have some type of *privileged access to them* — i.e., he must be in a privileged epistemic position with respect to them. But this raises a question: what type of privileged access can S have that is compatible with the constraints that Boghossian imposes on privileged self-knowledge? Since different notions of privileged access have been endorsed throughout history, we must examine them to determine which one plausibly underlies (A). Thus, let me consider as potential responses to the previous question some notions of privileged access that Alston (1971) has distinguished.

One important notion of privileged access that has played a crucial role in the history of philosophy is the one Alston calls 'infallibility'. According to authors that accept it, such as Descartes (1988, p.83) who wrote that 'I certainly seem to see, to hear, and to be warmed. That cannot be false (...)', we have privileged access to our thoughts because it is impossible for us to be wrong about them — i.e., because we are immune to error with respect to them. Though this notion of privileged access has been very influential, a brief reflection shows that it cannot be the notion underlying (A) because, if that were the case, (c) would be violated as Boghossian holds that our self-knowledge is fallible.

A second notion of privileged access that has also been extremely influential throughout history is what Alston calls 'omniscience' (though I use here the term 'transparency' to refer to it). The central idea behind this notion is this: if we enjoy transparency with respect to our own thoughts, it is impossible for us to ignore them when we are thinking them. Though some authors such as Locke (1975, p.335) have contended that we do enjoy this type of privileged access vis-à-vis some thoughts because 'it being impossible for anyone to perceive without perceiving that he perceives', this notion, which involves an immunity to ignorance, cannot be the one that underpins (A) insofar as Boghossian accepts that our self-knowledge is incomplete. Thus, if this notion of privileged access happened to underpin (A), (d) would be violated.

In contrast with infallibility and transparency, other authors have endorsed the view that we have privileged access to our thoughts since we are endowed with what

they refer to as 'indubitability' with respect to them. On this view, which has been endorsed by Hamilton (1878, p.188) who wrote that 'to doubt the existence of consciousness is impossible, for such a doubt could not exist, except in and through consciousness', our privileged epistemic position consists in the fact that it is impossible for us to doubt that we have some thoughts when we are thinking them. Even though this notion of privileged access has had historically a very wide acceptance since it affords an immunity to doubt, it cannot be the notion underlying (A) because, though the individual S is by stipulation neither aware of the shifts he undergoes nor a chemical expert, he is not unreflective (i.e., he can in principle entertain a global skeptical doubt). Thus, the notion of privileged access behind (A) cannot be indubitability under pain of violating (e).

A fourth notion of privileged access that Alston considers is what he labels 'incorrigibility'. The key idea behind it is this: if we are incorrigible with respect to our own thoughts, we are in a privileged epistemic position with respect to them since it is impossible for anybody to correct us about them. This position, which Ayer (1963, p.73) characterizes writing that 'the logic of these statements that a person makes about himself is such that, if others were to contradict him, we should not be entitled to say that they were right so long as he honestly maintained his stand', has attracted many as it affords an immunity to correction. However, a moment of brief reflection shows that it cannot be the notion of privileged access underlying (A) because someone observing S's travel (perhaps one of those who are responsible for transporting him) is in a position to correct him. In other terms, incorrigibility cannot underpin (A) without (f) being violated.

Alston distinguishes a fifth notion of privileged access, which he calls 'truth-sufficiency'. This notion of privileged access rests upon the following idea: if we have a true thought, it is impossible for us not to be justified in believing it since its truth is precisely that which justifies it.[12] In virtue of this, truth-sufficiency is, for Alston, a weaker analogue of transparency in the following sense: whereas an individual knows his thought if the thought in question is endowed with transparency only in virtue of the thought's occurrence, if his thought is rather endowed with truth-sufficiency, the thought's truth only guarantees its being justified, thus leaving open the possibility of ignorance. This notion of privileged access, which Shoemaker (1963, p.216) illustrates by mentioning that in certain cases '[...] being entitled to assert such a [first-person experience] statement does not consist in having established that the statement is true [...] but consists simply in the statement being true', is attractive since it is less demanding than transparence while still providing a type of epistemic immunity — i.e., an immunity to holding justified falsehoods, which leaves open the possibility of ignorance.

However, a closer look at this notion reveals that, despite its attractiveness, it cannot be at play in (A). Indeed, if two persons are endowed with truth-sufficiency

and think the same true thought, they would both in principle be equally justified in holding it, regardless of how much attention they pay to it. But this clashes with the constraint (g) that states that self-knowledge must be a cognitive achievement.

Finally, a sixth notion of privileged access that Alston presents is what he calls 'self-warrant'. The key idea underpinning this notion is this: in some cases, the mere fact that we believe something makes it impossible for us not to be justified in holding it since the very belief constitutes the justification. In light of this, Alston points out that self-warrant is a weaker analogue of infallibility in the following sense: while infallibility allows an individual to know his own thought just because he believes it, self-warrant allows an individual to have justification about his thoughts just in virtue of the fact that he holds them. James (1967, p.731) provides an illustration of this position when he considers the case of a man whose 'faith acts on the powers above him as a claim and creates its own verification.' This notion of privileged access involves, as in the other cases, a form of epistemic immunity — i.e., an immunity to having unjustified beliefs,[13] which leaves open the possibility of error.

Is this notion of privileged access the one that is at play in (A)? This seems to be the case in virtue of the fact that nothing that Boghossian says precludes it (in contrast with the other notions) and that it is substantially weaker than any of the other notions.[14] Since all the other plausible candidates have been ruled out, let us then suppose that (A) is true in virtue of S having self-warrant. Keeping this in mind, let us now consider (D), which states that if S knows his thought, he must be able to rule out the relevant alternative that he is thinking a water-thought. In virtue of this, considering that S must be capable of eliminating the relevant alternative, the notion of knowledge in (D) is not underpinned by self-warrant since self-warrant does not eliminate the possibility of error. A stronger notion of privileged access is required. But, if Boghossian appeals to a stronger notion of privileged access (e.g., infallibility), he equivocates between different notions of privileged self-knowledge that are underpinned by different types of privileged access.[15]

In virtue of this, Boghossian is then confronted to a dilemma: if the notion of privileged access underpinning (A) is not self-warrant but some other notion, he violates from the start the very conditions that he imposes on privileged self-knowledge (since privileged self-knowledge has to be, according to him, a cognitive achievement and also has to be consistent with the possibilities of error, ignorance, doubt and correction) and if the notion of privileged access underpinning (D) is self-warrant, the antecedent of (D) cannot be true in light of the constraint imposed by the consequent (i.e., the need to eliminate the relevant alternative to preclude the possibility of error). In order to make the antecedent of (D) true, Boghossian is then forced to adopt another notion of privileged self-knowledge grounded on a stronger notion of privileged access and, thus, equivocate. Considering this, the argument fails and the paradoxical conclusion can be successfully blocked.

5. Some Other Responses to the Argument

As I mentioned in the introduction, the slow-switching argument has been closely examined for over twenty-five years now, and it has elicited several responses that provide ways of resisting the conclusion. However, even though I agree with the spirit of many of these responses, I believe that the proposal that I have articulated previously is better than many others in a central respect. In order to see this, let me consider briefly some of the main previous responses that the argument has generated.

Falvey and Owens argue that the slow-switching argument founders in light of the fact that Boghossian fails to make a very important distinction. According to them (1994, p.109–10), the thesis that one's self-knowledge is privileged in the sense that it is authoritative and direct can be understood in two different senses: one can know the contents of one's own thoughts without relying on inferences from one's environment (which is a type of knowledge they call *introspective knowledge of content*) and one can know the contents of one's own thoughts by being able to determine without empirical investigation, for any two thoughts, whether they have the same content or not (which is a type of knowledge they call *introspective knowledge of comparative content*).

If this distinction is accepted, Falvey and Owens contend that one is in a good position to reject the argument arguing that, even if content externalism is at odds with S's introspective knowledge of comparative content, it is consistent with his introspective knowledge of content.[16] The main reason they give to justify this claim is this: in order for a relevant alternative to undermine S's claim that he has introspective knowledge of content of his own thought t, the relevant alternative must be such that, if it were true, S would still believe the same thought t. But this cannot be the case, according to them, as one can appreciate by reflecting on the case of Susan, who is stipulated to undergo experiences that are similar to those of S:

> Does it follow that if Susan were on Twin Earth thinking that twater is a liquid, she would still believe that she was thinking that water is a liquid? Certainly not — because the contents of a Twin-Earthian's second-order beliefs are determined by her environment just as the contents of the contents of her first-order beliefs are. If Susan were on Twin Earth, thinking with Twin-Earth concepts, she would not — indeed, she *could* not — believe that she was thinking that water is a liquid, anymore that she could think that water is a liquid. (1994, p.117, my emphasis)

In virtue of this, the response of Falvey and Owens consists in arguing that the initial assumption (A) is not undermined by S's failure to rule out relevant alternatives because this failure only precludes S's introspective knowledge of comparative

content and because the relevant alternative is no threat to S's introspective knowledge of content in virtue of the fact that S cannot think a water-thought when he is on Twin Earth.

Following Gibbons, Vahid (2003) introduces another way to resist the slow-switching argument. In particular, considering that Gibbons (1996, p.294) holds that the main lesson of the argument consists in showing that 'our knowledge of our own thoughts is more susceptible to empirical contingencies than we have believed', Vahid (2003, p.279) contends that there are two different ways in which we may interpret the susceptibility to empirical contingencies of our self-knowledge. We may construe this character in terms of a *positive* dependence of our beliefs on our environment (in which case our actual experience plays an appropriate role in producing the justification of our beliefs) or in terms of a *negative* dependence (in which case the justification of our beliefs would be undermined if our experience were different from what it actually is).

However, Vahid argues that, regardless of the case, the susceptibility of our self-knowledge to empirical contingencies in unproblematic. Indeed, in the case of positive dependence, Vahid maintains that, since advocates of privileged self-knowledge usually hold that S's belief that he thinks a water-thought stems from a process of reflective thought, no further evidence (and, specially, no empirical evidence) is needed to justify it. In light of this, the notion of dependence at play in the argument cannot be the positive one. And, in the case of the negative dependence, Vahid remarks that it does not threaten the privileged status of our self-knowledge because one may accept that our self-knowledge is privileged in the sense that its justification is not generated or maintained by actual empirical evidence while admitting that it would be defeasible by empirical evidence in counterfactual circumstances. Thus, since S' privileged self-knowledge is not precluded under any interpretation of the notion of susceptibility to empirical contingencies, Vahid concludes that the slow-switching argument founders.

Finally, Morvarid (2015) has recently developed a third approach to block the argument. The strategy of Morvarid relies upon an important insight initially articulated by Falvey and Owens. In order to be able to justify (D), Falvey and Owens (1994, p.116) note that Boghossian seems to rely on a certain epistemic principle that provides a necessary condition for knowledge:

(R) If (i) there is a relevant counterfactual situation in which q is true instead of p, and (ii) S cannot exclude the possibility that q is the case, then S does not know that p.

Now, Morvarid points out that this principle is ambiguously formulated to the extent that it is not clear how the condition 'S cannot exclude the possibility that q is the case' is to be interpreted. Considering this, Morvarid considers the three main

candidates that have been proposed in the literature to make sense of the idea that, in order to know his own thought, S has to be able to rule out the relevant alternative:

(R_1) If (i) q is a relevant alternative to p and (ii) S's belief that p is the case is based on evidence that is compatible with its being the case that q, then S does not know that p.

(R_2) If (i) q is a relevant alternative to p and (ii) S's justification for his belief that p is such that if q were true, then S would still believe that p, then S does not know that p.

(R_3) If (i) q is a relevant alternative to p and (ii) S cannot discriminate between the actual situation in which p is true and the counterfactual situation in which q is true, then S does not know that p.

After reviewing in detail all these candidates, Morvarid concludes that none of them is able (c_1) to provide plausible necessary conditions for knowledge and (c_2) to undermine the claim that S has privileged self-knowledge. In the case of R_2, Morvarid (2015, p.20) points out, echoing Falvey and Owens, that it does not undermine the claim that S has privileged self-knowledge in light of the fact that 'second-order beliefs inherit their contents from first-order thoughts.' In virtue of this, if we suppose that S is on Twin Earth thinking a twin thought, he simply cannot believe in that situation that he is thinking a water-thought. Considering this, as the antecedent of R_2 cannot be true, the claim that S has privileged self-knowledge is not undermined.

In the case of R_1, Morvarid argues that the principle is not a good candidate because it is subject to many counterexamples. In particular, Morvarid asks us to consider a situation where two indistinguishable twins (which he calls Trudy and Judy) live in Oscar's neighborhood and where Oscar sees Judy sitting on a chair. On the basis of the evidence that he has (which is perceptual),[17] Morvarid claims that it is clear that Oscar *knows* that *she* (=Judy) is sitting on a chair. But, as Morvarid (2015, p.24) remarks, 'the evidence on which Oscar's belief that p [which is the Russellian proposition ⟨Judy, the property of sitting in the chair⟩] is based is compatible with its being the case that q [which is the Russellian proposition ⟨Trudy, the property of sitting in the chair⟩]. For, by hypothesis, Judy and Trudy have the exactly similar superficial properties and Oscar acquires the same visual appearances in the actual and the counterfactual situation.' Morvarid then concludes that in this case, though the antecedent of R_1 is satisfied, R_1 is not a good candidate because it fails to give a plausible necessary condition on knowledge as it violates our intuitions.

Finally, in the case of R_3, Morvarid introduces another counterexample that provides good evidence to hold that R_3 fails to provide a plausible constraint on knowledge. He asks us to suppose that, when Oscar sees Judy sitting in the chair, he uses the definite description 'the girl who is sitting on the chair now' in order to intro-

duce a new name (e.g., 'Sara') that is fixed by the description. On the basis of this assumption, he then argues as follows:

> Intuitively, in such a circumstance, Oscar is in a position to know the conditional proposition that if the girl who is sitting on a chair exists then Sara is the girl who is sitting on a chair (call this proposition r) [...] Now, consider the relevant counterfactual situation in which Trudy, rather than Judy, is sitting on the chair. In this situation, clearly, proposition r is false and another conditional holds: if the girl sitting on the chair exists, then Trudy is the girl sitting on the chair (call this proposition s). Here, s is a relevant alternative to r, but Oscar cannot discriminate between the actual situation in which r is true and the counterfactual situation in which s is the case, as he cannot discriminate between Judy and Trudy. So, according to R_3, Oscar does not know that r, but this is highly counterintuitive. (2015, p.31–2)

Since no epistemic principle satisfies both (c_1) and (c_2), Morvarid concludes that the slow-switching argument does not get off the ground. Now, while I am sympathetic to all these proposals to block the argument, I believe they all exhibit a common shortcoming: none of them really attempts to address the main worry that motivated Boghossian initially to present the argument. Indeed, though Falvey and Owens suggest that what they refer to as introspective knowledge of content is immune to the argument, they are silent with respect to the conditions under which it obtains. The failure to address the worry is also rather conspicuous in the other authors insofar as Vahid (2003, p.386) writes in the last paragraph of his paper that a key problem remains open, which is 'to explain how privileged self-knowledge is possible if our concepts are environmentally determined', and insofar as Morvarid ignores the question by claiming that the slow-switching argument does not get off the ground in the first place.

In contrast with these responses, the proposal that I have articulated here enables us to provide at least a partial answer to Boghossian's question: given the conditions (a)–(g) that he imposes on privileged self-knowledge, the only circumstance in which, granting that content externalism is the case, we have privileged knowledge of our thoughts is when the type of privileged access that we have with respect to them is self-warrant.[18] The upshot of this claim is that, since self-warrant is a weak and deflated type of privileged access, the notion of privileged self-knowledge that one can defend in conjunction with semantic externalism is a weak and deflated notion that is very far removed from the immoderate views on self-knowledge endorsed by Descartes and others.[19]

6. Addressing Some Potential Objections

Though my proposal to resist Boghossian's argument suggests that there is a type of privileged self-knowledge that is compatible with semantic externalism, a number of

objections can be raised with respect to it. After rehearsing here the objections that I take to be most pressing, I present some responses. The first objection stems from the following observation: self-warrant guarantees only belief and justification on a systematic basis, but not truth. Consequently, since self-warrant does not preclude the possibility of error, how can a notion of privileged self-knowledge based on self-warrant be a form of genuine knowledge if truth is not systematically guaranteed?[20]

My response to this is that Boghossian himself concedes that our privileged self-knowledge is fallible — i.e., that it is consistent with the possibility of error. In light of this, our notion of privileged self-knowledge does not have to guarantee truth on a systematic basis to qualify as knowledge.[21] And, if some insist that our self-knowledge must be infallible in order to guarantee truth on a systematic basis, I respond that, if we accept this, we end up with a notion of privileged self-knowledge that is too stringent for human beings.

A second objection goes as follows: a weak and deflated type of privileged self-knowledge is unable to do any of the work that is required from it. In particular, one may raise, within this perspective, a more specific worry: how can we have a notion of intentional action that is robust enough if our notion of privileged self-knowledge based on self-warrant is so weak and deflated?[22]

In response to this worry, consider my intentional action of drinking water when I am thirsty. A brief reflection shows that carrying out the intentional action requires on my part having some beliefs (e.g., the belief that the glass cup in front of me is full of water) and some justification for them (e.g., the perceptual evidence that is generated by the cup and its contents), but not that these beliefs are systematically true. Indeed, intentional actions can fail for a variety of different reasons (e.g., the liquid in the cup turns out to be alcohol and not water), and our deflated notion of privileged self-knowledge, which allows the possibility of error, enables us to explain some of these failures.[23]

Finally, a third objection consists in claiming that we can provide a good account of the observations associated with privileged self-knowledge (e.g., deferring authority to others regarding knowledge claims they make about their own thoughts and expecting the same in return) without appealing to the idea that each individual has a privileged epistemic position vis-à-vis his thoughts.[24] According to this objection, which is draws its inspiration from certain remarks made by Wittgenstein, having privileged self-knowledge just involves accepting a particular language-game in which we defer authority to others regarding their knowledge claims about their own thoughts and reject as nonsense any claim of authority made by others vis-à-vis the assertions we make about our thoughts.[25]

My response to this objection is that, if having privileged self-knowledge just boils down to accept a certain language-game, having privileged self-knowledge no longer qualifies as a cognitive achievement. And this alternative appears to be im-

plausible because there are cases afoot wherein the knowledge that I have of my own thoughts is the result of an activity that I undertake (i.e., directing my attention to some thoughts) and not of anything that others do.

7. Conclusion

Let me recap. I have argued here that Boghossian's slow-switching argument is flawed because it trades on an equivocation between two different notions of privileged access that underpin the thesis of privileged self-knowledge. I have also shown that my response to the argument is better than other alternatives since it enables us to provide a partial answer to the question raised by Boghossian concerning the conditions under which privileged self-knowledge is possible given content externalism: it is only when the type of privileged access that we have with respect to our thoughts is self-warrant, as all the other options are precluded by Boghossian's constraints on self-knowledge.

The main upshot of my response consists then in showing that a weak and deflationary type of privileged self-knowledge is not inconsistent with content externalism and that, in light of this, one can in principle construct a general theory of mental content that integrates both theses. A couple of key tasks that remain then to be accomplished (and that I intend to pursue in future works) are the elaboration of such a theory and the development of a more systematic account of the notion of privileged self-knowledge sketched here that allows the possibilities of error, ignorance, correction, and doubt, and that is not cognitively insubstantial.

Acknowledgments

Earlier versions of this article were previously presented to and discussed with different audiences at the City College of New York, the CUNY-Graduate Center, Ohio State University, Denison University, Metropolitan State University of Denver and Universidad Autónoma Metropolitana-Iztapalapa. In addition to my colleagues at Metropolitan State University of Denver, I would like to thank for critical feedback and insightful questions Lourdes Valdivia, Carla Merino, Barbara Fultner, Alberto Cordero, David Weisman, Jorge Morales, Steve Vogel, Jonathan Maskit, Max Fernández, Godfrey Guillaumin, Carlos Romero, Rodrigo Campos and Esteban Withrington.

References

Alston. W. 1971. Varieties of Privileged Access. *American Philosophical Quarterly* **8**(3): 223–41.

Ayer, A. J. 1963. *The Concept of a Person and other Essays*. St. Martin's Press: New York.

Bernecker, S. 2004. Memory and externalism. *Philosophy and Phenomenological Research* **69**(3): 605–32.
Boghossian, P. A. 1989. Content and Self-Knowledge. *Philosophical Topics* **27**(1): 5–26.
———. 1997. What the Externalist can know A Priori, *Proceedings of the Aristotelian Society* **97**: 161–75.
Burge, T. 1979. Individualism and the Mental. *Midwest Studies in Philosophy* **4**(1): 73–122.
———. 1988. Individualism and Self-Knowledge. *Journal of Philosophy* **87**(11): 649–63.
Davidson, D. 1984. *Inquiries into Truth and Interpretation*. New York: Oxford University Press.
———. 1987. Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association* **60**(3): 441–58.
Descartes, R. 1988. *Selected Philosophical Writings*. Trans. by J. Cottingham, R. Stoothoff and D. Murdoch. New York: Cambridge University Press.
Falvey, K.; Owens, J. 1994. Externalism, Self-Knowledge and Skepticism. *Philosophical Review* **103**(1): 107–37.
Hamilton, W. 1878. *Lectures on Metaphysics and Logic. Vol.I. Metaphysics*. New York: Sheldon and Company.
Gibbons, J. 1996. Externalism and knowledge of content. *Philosophical Review* **105**(3): 287–310.
Heil, J. 1988. Privileged access. *Mind* **97**(1): 238–51.
James, W. 1967. *The Writings of William James. A Comprehensive Edition*. Ed. by J. McDermott. New York: Random House
Locke, J. 1975. *An Essay on Human Understanding*. Ed by P. H. Nidditch. Oxford: Clarendon Press
Mendola, J. 2008. *Anti-Externalism*. New York: Oxford University Press
McKinsey, M. 1991. Anti-Individualism and Privileged Access. *Analysis* **51**(1): 9–16.
Morvarid, M. 2015. The epistemological bases of the slow-switching argument. *European Journal of Philosophy* **23**(1): 17–38.
Putnam, H. 1975. The Meaning of "Meaning". *Minnesota Studies in the Philosophy of Science* **7**: 131–93.
———. 1981. *Reason, Truth and History*. New York: Cambridge University Press.
Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson & Co.
Searle, J. R. 1983. *Intentionality. An Essay in the Philosophy of Mind*. New York: Cambridge University Press.
Shoemaker, S. 1963. *Self-Knowledge and Self-Identity*. Ithaca, NY: Cornell University Press.
Tanney, J. 1996. A Constructivist Picture of Self-Knowledge. *Philosophy* **71**: 405–22.
Vahid, H. 2003. Externalism, Slow-Switching and Privileged Self-Knowledge. *Philosophy and Phenomenological Research* **66**(2): 370–88.
Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Basil Blackwell.

SERGIO ARMANDO GALLEGOS
Philosophy Department
Metropolitan State University of Denver
CN 307J
Campus Box 49, P.O. Box 173362
Denver, CO 80217-3362
sgalle36@msudenver.edu

Notes

[1] In particular, Putnam (1975) proposes a natural kind externalism according to which the meanings of natural kind concepts such as *water* or *aluminum* are individuated by the substances that individuals causally interact with while Burge (1979) advocates a social externalism in which the meanings of concepts such as *arthritis* are individuated on the basis of the linguistic conventions and the social practices of the group to which they belong

[2] For instance, Davidson (1987, p.441) writes the following: 'It is seldom the case the case that I need or appeal to evidence or observation in order to find out what I believe; normally I know what I think before speak or act.' It is important to notice here that Davidson uses the adverb 'seldom' instead of 'never' since this shows that, for him, our privileged self-knowledge is not infallible.

[3] For instance, see Mendola (2008, p.2–3): 'The rain on the street seems to help constitute my seeing it, so that if something else were there in the place of rain, that other thing and my internal state of openness would constitute my seeing something else.'

[4] The philosophical arguments that motivate semantic externalism appeal to thought experiments involving individuals traveling to a *Doppelgänger* planet dubbed Twin Earth (Putnam 1975 and Burge 1988) or the creation of a physical duplicate of some individual ('Swampman') that shares with the original all of his intrinsic features but not his history or external relations (Davidson, 1987).

[5] This example is from Ryle (1949, p.160)

[6] The issue that Putnam presents in this passage concerns terms such as 'tree', but a similar problem arises with respect to concepts.

[7] Boghossian (1989, p.6): 'I have to confess, however, that at the present time I am unable to see what these conditions might be.'

[8] In a parenthetical remark, Boghossian acknowledges the existence of complexities induced by Gettier counterexamples, but leaves them aside.

[9] There are many illustrations of the fact that privileged self-knowledge requires effort and introspective focus from us. Jane Austen presents one of the best known examples in the following passage from *Emma* cited by Julia Tanney (1996, p.406): 'Emma's eyes were instantly withdrawn; and she sat silently meditating in a fixed attitude for a few minutes. A few minutes were sufficient for making her acquainted with her own heart. A mind like her, once opening to suspicion, made rapid progress. She touched — she admitted — she acknowledged the whole truth. Why was it so much the worse that Harriet should be in love with Mr. Knightley than with Mr. Churchill? Why was the evil so dreadfully increased by Harriet's having some hope of return? It darted through her, with the speed of an arrow, that Mr. Knightley must marry no-one but herself.'

[10] In this respect, despite disagreements on other several issues, Burge and Boghossian are in agreement since they both think that giving up the thesis that we have privileged self-knowledge is absurd.

[11] In virtue of this, the slow-switching argument is clearly different from the argument that Boghossian puts forth in (1997), which is explicitly a *reductio*: the upshot of that argument, as opposed to the intended upshot of the slow-switching argument, is to suggest that at least one of its premises is false.

[12] Alston (1971, p.234) characterizes the relation of truth-sufficiency to knowledge in the

following terms: 'Knowledge involving truth-sufficiency is a sort of limiting case of direct knowledge, for here what is taken to justify the belief is something that is independently required for knowledge, viz., the truth of the belief. Thus, nothing over and above the other two conditions for knowledge [truth and belief] is required for the satisfaction of condition B [i.e., justification].'

[13] For Alston (1971, p.235), although truth-sufficiency and self-warrant are very similar in a certain respect, the latter involves a different type of epistemic privilege: 'Whether I enjoy self-warrant or truth-sufficiency (or both) vis-à-vis my current thoughts and feelings, it will follow in either case that whenever I have a true belief to the effect that I am thinking of feeling x at the moment, I can correctly be said to know that I am thinking or feeling x. (...) However, they carry different implications as to what can said short of a full knowledge claim. Enjoying self-warrant in this area guarantees that any belief of this sort is justified. It protects against the possibility of unjustified belief formation. Whereas truth-sufficiency makes no such guarantee; it is compatible with the existence of some unjustified beliefs in the appropriate range.'

[14] Alston (1971, p.236) shows a strong preference for self-warrant *qua* form of privileged access in virtue of the following advantages that it offers: 'It escapes the objections urged against claims of infallibility, omniscience, indubitability and incorrigibility. It allows for cases in which a person mistaken about his current mental states, (and of course it puts no limit at all on the extent to which a person may be ignorant of his current mental states), and it even allows for cases in which someone else can show that one is mistaken. And, at the same time, it specifies a very definite respect in which a person is in a superior epistemic position *vis-à-vis* his own mental states'

[15] The objection that I am raising here to the slow-switching argument is very similar to an objection raised by Bernecker (2004) to the 'memory argument', insofar as both involve diagnosing an equivocation.

[16] Falvey & Owens (1994, p.118): 'It follows that the externalist can safely endorse the relevant alternatives (...) quite generally, without opening herself to the charge that the externalism she espouses undermines introspective knowledge of content.'

[17] Morvarid (2015, p.25) points out that there are two different ways to understand the notion of perceptual evidence: one may construe it in terms of 'a set of perceptual appearances' or in terms of 'a contentful state whose content is determined by the environment'. However, regardless of the notion of perceptual evidence that one has in mind, he shows that one can construct counterexamples that undermine the candidates for the role of epistemic principle.

[18] This is just one necessary condition. There are broader questions concerning whether there are other necessary conditions (in addition to those established by the criteria (a)–(g)) and which are the kinds of thoughts that we can have privileged knowledge of on the basis of self-warrant that, given space limitations, I cannot address here.

[19] Consequently, the notion of privileged self-knowledge that is consistent with content externalism cannot be used, in virtue of its weak and deflated character, for the kind of foundationalist project undertaken by Descartes. But it is robust enough to vindicate the truism stated by Davidson in footnote 2.

[20] I thank Carla Merino for putting forth this objection in conversation. What follows is an attempt to answer it.

[21] A good illustration of this view is provided by Davidson (1984, p.136) who, though accept-

ing the importance of truth as the central element that is contributed by terms to thought, nevertheless concedes in some early texts the necessity to account for the possibility of error in a way that does not destroy knowledge when he writes the following: '(. . .) once the theory begins to take shape, it makes sense to accept intelligible error and to make allowance for the relative likelihood of different kinds of mistake.' For further discussion, see also Mendola (2008, p.266).

[22] I am grateful to Lourdes Valdivia for presenting to me this objection in conversation.

[23] In this case, the quenching of my thirst is part of the conditions of satisfaction of my intentional drinking water from the cup but, since the cup does not contain water, my intentional action of drinking water from the cup fails. For further discussion of similar cases, see Searle (1983, p.80–3).

[24] I thank Barbara Fultner for putting forth this objection during the course of a Q&A session.

[25] The passages that suggest this interpretation are the following ones: 'It can't be said from me at all (except perhaps as a joke) that I *know* I am in pain, except perhaps that I *am* in pain? What is it supposed to mean — except perhaps that I *am* in pain? Other people cannot be said to learn from my sensation only from my behavior, — for I cannot be said to learn of *them*. I have them. The truth is: it makes sense to say about other people that they doubt whether I am in pain; but not to say it about myself.' (1953, §246) and "Our mistake is to look for an explanation where we ought to look at what happens as a 'proto-phenomenon'. That is, where we ought have said: that language game is played."(1953, §654)