

Semantic Externalism, Authoritative Self-Knowledge,  
and Adaptation to Slow Switching

Andrew F. Smith  
Department of Philosophy  
Stony Brook University  
Stony Brook, NY 11794  
afsmith@ic.sunysb.edu  
309-530-3726

There has been considerable debate in recent years by those concerned with issues surrounding the doctrine of semantic externalism, i.e., that the contents of at least some of our thoughts are individuated according to the physical and/or social environment, over whether it is compatible with an adequate account of authoritative self-knowledge.<sup>1</sup> Of particular interest to those struggling with the epistemological consequences of social externalism is Peter Ludlow's (1995) modification of Paul Boghossian's (1989) argument against compatibility made in his influential article 'Content and Self-Knowledge'. Ludlow provides this modification in order to circumvent the objection made by Ted Warfield (1992) that Boghossian's argument is in fact invalid and that his so-called 'Incompatibility Thesis' (as Warfield refers to it; hereafter IT) is therefore on tenuous footing. So modified, if Ludlow's own argument holds up, the core features of Boghossian's original claim are rescued and so too is his argument against compatibility. We are thus left, as the now familiar line of argumentation goes, in the position of having to draw one of two conclusions: either we must give up the idea that we do indeed maintain authoritative self-knowledge or we must reject semantic externalism.

In this essay, however, I intend to show that while Ludlow's modification may indeed make Boghossian's argument valid, the modification is itself highly problematic. In particular, it is unable to cope with certain key difficulties that become apparent through a more thorough examination of the social externalist position laid out by Ludlow himself. We shall therefore see that his attempt to establish the incompatibility of externalism and authoritative self-knowledge by means of this modification proves unsuccessful. To anticipate, by more closely examining the logic of certain of the presuppositions behind Ludlow's argument and drawing on some fairly straightforward intuitions that arise in the

process, it becomes clear that a plausible case for compatibility can be made that throws the viability of IT into doubt. In order to properly develop this argument, I proceed in the following manner: In Part I, I briefly lay out Boghossian's, Warfield's, and Ludlow's arguments, respectively. I shall then be in a position to show in Part II that Ludlow's modification of IT does not achieve its intended aim.<sup>2</sup>

## I

Boghossian bases his argument against compatibility upon the well-known slow-switching thought experiment proposed by Hilary Putnam (1975)—and updated by Tyler Burge (1979) to account for the relevance of social factors—in which an agent is unwittingly transported between Earth and Twin Earth.<sup>3</sup> Let us begin, Boghossian asserts, by assuming the truth of externalism. Let us then assume that slow switching does indeed occur and, furthermore, that when one undergoes slow switching he/she likewise undergoes the wholesale displacement of his/her Earth concepts by Twin Earth concepts (and vice versa when transported back to Earth). That is to say, all of his/her Earth concepts are displaced by coinciding Twin Earth concepts.<sup>4</sup>

Earth and Twin Earth are identical, so the thought experiment goes, save in one regard: while H<sub>2</sub>O fills the rivers and lakes on Earth, the rivers and lakes on Twin Earth are filled with a superficially and phenomenologically indistinguishable substance with the chemical make-up not of H<sub>2</sub>O but of XYZ. Keeping in mind that according to the externalist position our concepts are individuated according to the physical and/or social environment—i.e., that they are widely individuated—it must be the case that while our thoughts on Earth about H<sub>2</sub>O are about water, our thoughts on Twin Earth about XYZ must

instead be about twin water (hereafter twater). Given these conditions, Boghossian therefore asserts, it is entirely possible that a person S who has unwittingly undergone slow switching may know the contents of her thoughts at time t1, forget nothing, but at some later time t2 be unable to authoritatively say what she knew at t1.

This is, to be sure, a serious concern for the proponent of semantic externalism. For, as Boghossian states:

is it not precisely knowledge of this form—knowledge of what one has thought immediately after one has thought it—that we think of as central to our capacity for self-knowledge? We are struck by our ability to know, non-inferentially and authoritatively, that a certain mental event has occurred, immediately on its having occurred. (22)

Moreover, while the Twin Earth thought experiment is, of course, purely hypothetical, he argues that ‘it seems fairly easy’ (13) to describe other less hypothetical scenarios involving slow switching. Yet, in *every* such case, if we maintain the truth of externalism we encounter the very troubling conclusion that S is unable to have authoritative knowledge of the contents of her thoughts without first investigating her environment. And this ability of ours to have direct access to the contents of our thoughts, an ability that it seems strikingly clear we do possess, becomes difficult to explain.

Boghossian rightly asserts that one possible explanation, memory failure on S’s part, hardly gets to the point of the matter. In discussing the epistemological consequences of our mental content being widely individuated, ‘we ought to be able to exclude memory failure by stipulation. It is not as if thoughts with widely individuated contents might be easily known but difficult to remember’ (23). Boghossian thus advances the only

conclusion he finds left open to him: Strange as it may seem, S must *never have known* the contents of her thoughts in the first place. Insofar as this is the case, IT is verified. Either the externalist position or the authority of our self-knowledge must be abandoned.

According to Warfield, however, Boghossian is too hasty in drawing this conclusion. The compatibilist position simply is not as tenuous as he makes it out to be. Warfield points out that on his way to purportedly establishing the truth of IT, Boghossian provides (as he must) an argument for the incompatibility of externalism with a thinker knowing the contents of his/her thoughts via introspection. This more specific argument is the first—and arguably the most important—building block for IT. Hence, Warfield argues, ‘Boghossian’s argument for IT can be no better than his argument for the incompatibility of externalism and introspective self-knowledge’ (233). If it can be shown that the latter argument is invalid, it will become clear that the grounds for accepting the former are quite tenuous. Warfield therefore sets his sights squarely upon establishing that Boghossian’s argument for the incompatibility of externalism and self-knowledge via introspection is indeed invalid.

Warfield formalizes the argument provided by Boghossian as follows:

- (P1) To know that P by introspection, S must be able to introspectively discriminate P from all relevant alternatives of P.
- (P2) S cannot introspectively discriminate water thoughts from twater thoughts.
- (P3) If the Switching Case is actual, then twater thoughts are relevant alternatives of water thoughts.
- (C1) Therefore, S doesn’t know that P by introspection.<sup>5</sup> (234-235)

This conclusion, Warfield asserts, does not follow from the premises. All that follows is this weaker conclusion:

(C1') If the Switching Case is actual, then S doesn't know that P by introspection. Even if the first two premises are true, what is required of Boghossian to establish (C1) is to establish that the Switching Case—i.e., the proxy within the argument for real-world scenarios resembling the Twin Earth thought experiment—is *actual*.<sup>6</sup> Yet, Boghossian has established nothing of the sort. Although he has shown that being an unwitting victim of slow switching is certainly logically possible, he provides no basis to accept that slow switching actually occurs and that twater thoughts (or any other such twin thoughts, for that matter) are therefore a relevant alternative to water thoughts that require being ruled out before any claim to authoritative self-knowledge can be made. Thus, Warfield argues, Boghossian's argument for the incompatibility of externalism and self-knowledge via introspection is invalid and, consequently, IT proves questionable. It does not approach the more important question of the compatibility or lack thereof of externalism and authoritative self-knowledge; and it provides no basis for abandoning the compatibilist position.

Although he seems to agree in spirit with Warfield's compatibilism, Ludlow nonetheless seeks to exonerate Boghossian by showing that a modified version of his argument for IT circumvents Warfield's objection. On Ludlow's account, switching cases being actual—as premise (P3) suggests—does not cut to the heart of the matter. Demanding as much, he contends, in fact 'understates the conditions under which there are relevant alternatives to water thoughts' (46), for it is the *prevalence* and not merely the actuality of such cases that makes them relevant. And slow switching is in fact prevalent.

It occurs all the time as we move from one linguistic community to another—from home to work to social and community functions, and so on. Ludlow thus reformulates Boghossian's argument by revising premise (P3) and adding a fourth premise that, he argues, together make (C1) valid:

(P3') If switching cases in general are prevalent, then there are relevant alternatives to water thoughts.

(P4) Switching cases, in general, are prevalent. (46)

In proposing within premise (P3') that the prevalence of switching cases entails that there are relevant alternatives to water thoughts, alternatives that must be ruled out if one is to be warranted in claiming that he/she maintains authoritative self-knowledge, what Ludlow is insinuating more generally is that unwitting shifts in mental content are themselves a relevant consideration. This being so, given the truth of semantic externalism, it is plausible to conclude that as we travel between different circles of acquaintances the contents of our thoughts may well tend to shift. Given this state of affairs, moreover, it does seem that we are at risk—when we slide between linguistic communities—of not knowing the contents of these thoughts before we investigate our external environment.

Ludlow provides the example of Biff, a philosophy professor who is cognizant of the fact that when he engages with various groups of acquaintances—members of his own department and discipline, members of other departments and disciplines, his family, his friends outside of the academy—he effectively shifts linguistic communities (47). Now, Biff knows little of classical American philosophy and thereby has only partial knowledge of the term 'pragmatist'. He may be unaware that the term has different meanings in two or more linguistic communities in which he circulates. Insofar as Biff must defer to other

members of these various communities concerning the meaning of the term, he unfortunately runs the risk of not having direct and privileged access to the contents of his own thoughts. But, Ludlow warns:

this isn't just a story about hapless Biff. We routinely move between social groups and institutions, and in many cases shifts in the content of our thoughts will not be detected by us.... Nor are these cases even limited to obvious cases of movement. It may occur when we routinely cross campus to talk to colleagues in physics or psychology, or even when we pay routine visits to our favorite restaurant. (48)

Given this state of affairs, the compatibility of externalism and authoritative self-knowledge does once again appear to be compromised, and on grounds that Warfield's objection cannot combat. Establishing that slow switching between linguistic communities is prevalent would indeed seem to suffice likewise to establish the actuality of such switches. This being so, (C1) goes through and IT is given new life.

However, in the remainder of this essay we shall see that the compatibilist has an answer to Ludlow's claim—namely that slow switching cases the likes of which he describes, although arguably prevalent, have fleeting effects. It is quite plausible that those who undergo these switches easily adjusted to them; hence, they do not undermine authoritative self-knowledge after all. What this means, more concretely, is that premise (P3') proves to be highly suspect, if not altogether false. The prevalence of slow switching between linguistic communities does not necessarily entail that unwitting shifts in the contents of our thoughts actually can occur. As such, Ludlow's modification of Boghossian's argument is itself rendered dubitable.

## II

I do agree with Ludlow that the more plausible and interesting cases of slow switching are those that involve the occurrence of switching between linguistic communities. Most of us do routinely travel between distinctive linguistic communities and social circles; and, assuming the truth of social externalism, the meanings of any number of the terms we employ may tend to shift on a regular basis. However, do we regularly pass into new and unfamiliar communities—i.e., experience what I shall call *non-routine* switches? For instance, do we often relocate to unfamiliar locales or acquire new groups of friends and discard old ones? Not most of us. To be sure, Ludlow insinuates that the sort of switching cases upon which his argument focuses are those that are *routine* in nature—visiting acquaintances in other departments across campus, going to work and returning home, going out to one’s favorite restaurant, and so on.

The crucial point here is that although individuals may frequently experience cases of routine switching the likes of which Ludlow describes, cases of non-routine switching occur quite infrequently. This being so, I submit that non-routine switches should not be counted among the cases of slow switching considered by Ludlow to be prevalent. The burden of proof for the truth or falsity of premise (P3') should be placed squarely upon the relevance of the shifts in mental content caused by routine switches, insofar as it is they that make up the vast majority of the switches between linguistic communities that we undergo. (We shall see later that there is good reason to accept that the shifts in mental content caused by non-routine switches do not occur unwittingly anyhow.) We thereby can exclude non-routine switches from consideration as a factor within Ludlow’s argument

and henceforth concern ourselves solely with the relevance of shifts in mental content caused by routine switching.<sup>7</sup>

Let us now recall the case of Biff, our ostensibly hapless philosophy professor who has only partial knowledge of the meaning of the term ‘pragmatist’ and, as such, must defer to other members of the various linguistic communities in which he circulates for the individuating conditions of the term. Insofar as Biff routinely shifts between linguistic communities, so the story goes, he unwittingly falls victim to slow switching.

Or does he? If such switches go wholly undetected by Biff, it is certainly likely that he would not be aware of differences in the meaning of certain of the terms he employs and, by consequence, of shifts in the contents of his thoughts. It is not so clear, however (and here I point to the first of our ‘fairly straightforward intuitions’), that routine switches in fact are undetectable. By their very nature, it seems that routine switches must occur with some sense of regularity and according to some pattern, rough as it may be, that one has established for oneself. As such, it is quite plausible that Biff rather easily grows accustomed to such switches—that, in general, he does not fall victim to slow switching but instead comes to smoothly transition between the various linguistic communities he frequents. That is to say, Biff regularly *adapts* to routine slow switching between linguistic communities. Moreover, he likewise effectively adapts to the shifts in mental content that accompany them. We can, and arguably should, accept premise (P4) yet nonetheless assert that twin thoughts do not occur unwittingly.<sup>8</sup>

We must, of course, now ask the obvious question: How does this adaptation occur? To begin to develop a response, let us set the issue of routine switching aside for the moment and examine another feature of Ludlow’s argument. According to Ludlow, it

is Biff's limited knowledge of classical American philosophy that necessitates that he defer to other members of the various communities in which he circulates when the term 'pragmatist' is employed. And we have already seen that it is Biff's partial knowledge and subsequent deference that engenders the risk, when undergoing slow switching, of his being an unwitting victim of shifts in mental content. It is important to note that Ludlow seems to assume, as the consistent semantic externalist should, that Biff's having partial knowledge entails that he maintains what Steven Davis (2000) refers to as a 'conditional disposition' to defer. Biff is willing to alter his beliefs about the meaning of 'pragmatist' in light of considerations that reveal that these beliefs are mistaken.<sup>9</sup> Moreover, following Burge, Ludlow indicates that such deference is both conscious and voluntary.<sup>10</sup> Biff, it thereby would seem, is generally cognizant of when his knowledge is partial—that is to say, of when he does not maintain epistemic authority—and on such occasions he is equally aware that he must defer to those who apparently have more authoritative knowledge on the matter being scrutinized than does he.<sup>11</sup>

There is no reason to believe that Biff is unaware that differences in the meaning and usage of certain of the terms he employs can exist. His experience of shifts in the meanings of terms of which he has substantial knowledge, when undergoing switches, would suffice to inform him of this. This being so, there is every reason to accept that he is well aware that shifts in meaning can occur when traveling between distinctive linguistic communities. And insofar as he realizes that there are occasions on which his limited knowledge necessitates his deference, it is not hard to believe (and here I note a second intuition) that he is cognizant of the need to remain especially alert on those occasions to the possibility of shifts in the meanings of certain of the terms he employs. To put it

simply, it is arguable that Biff is *predisposed*, when he is aware of the need to defer, to being vigilant to differences in meaning. Seeking to acquire greater knowledge of the various concepts regarding which he has partial understanding—or, at the very least, having the capacity to follow the flow of conversation—would seem to require no less.

We can therefore readily regard Biff as being interested in and cognitively directed toward being aware of shifts in the meaning of terms between linguistic communities. He is not merely a feather in the wind, battered about by switches between linguistic communities. His long-habituated inclination to be able to effectively converse with others and to *not* be duped by shifts in the meanings of certain of the terms he uses provide adequate incentive to be vigilant to possible shifts in meaning—and thus, whether he is cognizant of it or not, to the shifts in mental content that accompany them.

Let us now return to the issue of routine switching. I argued above that the vast majority of the switches that Biff undergoes are familiar and at least roughly patterned. But, as Ludlow himself states, this certainly is not just a story about Biff, our humble philosophy professor. Ludlow is quite right, as premise (P4) suggests, that we grow up in a world in which routine switching between linguistic communities is prevalent: the norm rather than the exception.<sup>12</sup> Yet, we clearly can and do make sense of our world. We do develop the ability to understand the meanings of the terms we use, including in many cases different meanings for the same term. From the perspective that Ludlow has given us, then, the truly pertinent question seems to be not whether we can maintain authoritative self-knowledge while undergoing prevalent, routine slow switching. The better question is this: Why are we *not* battered about by these switches and the shifts in mental content that they can engender?

Given their nature, we generally become quite accustomed to recognizing not only that routine switches occur but also *when* they occur. That is to say, our (at least) roughly patterned exposure to routine switches between linguistic communities allows us (enter here a third intuition) to become adept at differentiating between these communities and thus at recognizing their boundaries. We are able to identify our encounters by means of what sociologist John Armstrong (1994) calls ‘linguistic border guards’ (144): semantic signals warning us of when we have crossed linguistic boundaries.<sup>13</sup> The very fact that we can rather easily acknowledge that cases of slow switching are prevalent and are able upon reflection to stipulate the various linguistic communities in which we circulate, as Ludlow does, is itself a testament to this. Furthermore, it stands to reason that, like Biff, most of us will tend to be more vigilant to shifts in the meanings of certain of the terms we employ when (1) we have deferred and (2) we are aware that we have switched linguistic communities. We have already seen that Ludlow rightly assumes we are cognizant of when we defer. We now likewise have established that routine switches between linguistic communities are readily detectable. It therefore is reasonable to conclude that we are indeed generally alerted to those occasions on which our deference requires heightened vigilance to shifts in the meanings of certain of the terms we employ—specifically, those regarding which our knowledge is (necessarily self-reflexively) partial. In sum, that is, our cognizance of (a) deferring, (b) *a fortiori* our partial knowledge of those terms regarding which we defer, and (c) occurrences of a routine switching in combination quite naturally contribute to what we might say is our conditional disposition to be vigilant to shifts in meaning of the terms we employ.

But this is not the end of the story, for it is not merely the case that we aim to become increasingly aware of such shifts. If I may be permitted to introduce one final intuition, we likewise *actively, interestedly*, in many cases *fervently* seek to adapt to these shifts. Whether we are always fully conscious of it or not, we seek (or, with Burge, certainly should seek) to increase our knowledge by coming to understand how the utterances of others convey new concepts and ideas, or at the very least to recognize the limitations of that concerning which we ourselves can authoritatively speak. Maybe this adaptation occurs out of an inherent desire to increase our knowledge. Or maybe it simply occurs in order that we can avoid the public embarrassment that may arise by using a term or phrase incorrectly or fail to track the flow of conversation. In any case, this active adaptational process is arguably a normal feature of human life. Although the vindication of such a claim is ultimately contingent upon empirical investigation, it seems fair to speculate that initial deference concerning those terms we use regarding which we recognize our knowledge to be partial and the subsequent acquisition of greater knowledge of their multifarious meanings—or at least greater understanding of the limitations of our epistemic authority—are simply natural aspects of human experience. Regardless, the theoretical plausibility hereof raises serious questions about the viability of IT as it has been defended by Ludlow.

Before proceeding to our conclusion, it is important to reply to a possible objection to the argument I have here presented. The objection runs roughly as follows: It would seem that the awareness by an agent of the need to be vigilant must be a mental state *M* with a given content *C*. *M* and its content *C* maintain a sustained presence in an agent who undergoes slow switching; this has to be so if he/she is to be able to remain vigilant

throughout the duration of the switch to shifts in the meanings of terms. As such, the role of M is likewise absolutely crucial to adjusting to shifts in mental content and, henceforth, to maintaining authoritative self-knowledge. If externalism is true, however, M cannot enjoy sustained existence through slow switching. No mental state can. Thus, either we must hold onto the sustenance of M and give up the truth of externalism or admit the transitory nature of M and give up authoritative self-knowledge. We once again end up with incompatibility.

Does my argument depend upon the sustenance of M and therefore attempt to solve the problem of compatibility in a rather surreptitious manner? No, it does not.<sup>14</sup> In accepting the truth of externalism, I give up any guarantee to such sustenance of M. But there *is* nonetheless something that is preserved through these switches whereby our awareness of the need to be vigilant is maintained: memory. Recall Boghossian's assertion that memory failure does not adequately address the ability of S not to know at t<sub>2</sub> what she knew at t<sub>1</sub>. There is no reason at all to suppose that memory failure on S's part has occurred. Moreover, as Burge (1998) has pointed out, memory in fact has preservative powers that extend through a switch. While our mental states may well be compromised by slow switching, the contents of our memories are not. On Burge's account, we thereby can have authoritative non-empirically warranted self-attribitional knowledge of the form *I think (believe, judge) that p*, where *that p* includes some relevant concept R.<sup>15</sup> This being so, one could have a thought like *I believe that vigilance to a shift in mental content is warranted* both before and after a switch. The content of the thought is fixed through memory, not by the external environment in this case. Memory therefore acts as a 'link' between the content of earlier thoughts and a current, memory-induced thought. As Burge

himself states, 'Preservative memory normally retains the content and attitude commitments of earlier thinking, through causal connections to the past thinkings. That is one of its functions—maintaining and preserving a point of view through time' (357). Awareness need not depend upon reference to a particular mental state but instead upon non-empirically warranted self-attribitional knowledge: knowledge that is sustained through switches due the preservative power of memory.

The persistent objector could, of course, now contend that such recourse to memory provides an easier way to derail Boghossian's argument than I here present. His argument that S does not know the content of her initial water thought when she reflects at  $t_2$  can now perhaps be overcome straightaway. That is, perhaps S does know after moving from Earth to Twin Earth that she was thinking that water is wet, being able to do so thanks to the preservative power of her memory. This is the strategy taken by Burge. Yet, I believe that Boghossian has a possible rejoinder that necessitates taking the route I have to thwart IT. Might it not be the case, he could argue, that memory is shifty when we are not employing memory in its linking function. He could grant that when it is so employed—when S directly wonders whether 'this' content is the same as 'that' content—she knows at  $t_2$  what she knew at  $t_1$ . But are there not considerable periods of time in which the linking function of memory is not being employed? Maybe during these periods memory is just as shifty as are mental states. There could very conceivably be a time  $t_n$  at which the linking function is not being employed and therefore at which S does not in fact have authoritative self-knowledge.

Boghossian's rejoinder puts Burge in a tough position, one to which I believe my argument does not succumb. For we can grant the shiftiness of memory as Boghossian

might describe it but *still* maintain that this need not necessitate the endorsement of IT. To avoid falling prey to Boghossian's rejoinder, it would seem that Burge would have to maintain that the linking function of memory is perpetually employed. But we certainly are not perpetually employing our memories in this manner, nor should we need to in order to maintain authoritative self-knowledge of the contents of our thoughts. To be sure, the linking function of memory does suitably guarantee authoritative self-knowledge when it is so employed. Boghossian's rejoinder does not put this in jeopardy. Yet, if my argument holds up, we can now accept (*pace* Burge) that we actually only need to employ memory in its linking function when deferring. It is only under this condition, on Ludlow's account, that we can fall victim to undetected shifts in the contents of our thoughts. Thus can we grant that memory itself is shifty in a way not allowed for by Burge without having to give up the claim to the plausibility of compatibility.

This is not to say that one cannot be caught off guard on occasion by shifts in meaning under routine conditions. Unfamiliar yet authoritative interlocutors who connote idiosyncratic meanings upon certain terms may be introduced into otherwise familiar linguistic communities from time to time, and adaptation to the shifts in mental content that their utterances engender may not necessarily be automatic. But even though routine switching is prevalent, such occurrences surely are not. Such occurrences, in fact, are quite rare—so rare that they hardly can be seen to rise to the level of being legitimate threats to the general stability of our self-knowledge. It thereby is certainly plausible that privileged access to the contents of our thoughts is *not* compromised due to the prevalence of routine slow switching between linguistic communities. The epistemic effects of such switches

are easily adjusted for insofar as we are predisposed to vigilance and, consequently, actively pursue adaptation.

This of course leaves room to base an argument for incompatibility upon cases of non-routine switching. However, it stands to reason that given the unfamiliar nature of new linguistic communities, we are even more aware of our need to be vigilant to shifts in the contents of our thoughts when we must defer to their members. That is to say, non-routine switches are *phenomenologically distinguishable* from routine switches. Thrown out of our routine, we quite easily recognize when we cross borders into new communities and are therefore generally all the more alert to the need to be vigilant to the potential shifts in meaning that accompany these border crossings. That non-routine switches occur does not compromise the argument I here present any more than does the occurrence of routine switches.

Let us now finally turn our attention squarely upon premise (P3') of Ludlow's modification. We have excluded from consideration cases of non-routine switching insofar as they do not count as prevalent (and we have subsequently seen that they are detectable anyhow). Moreover, given that we have good reason to accept that routine switches are themselves readily detectable by us, we can easily adjust to the corresponding shifts in the contents of our thoughts engendered by them. As such, the prevalence of slow switching does not in fact necessarily entail the occurrence of unwitting shifts in mental content. Premise (P3') thereby is rendered suspect, if not false outright, and the viability of Ludlow's modification of Boghossian's argument against the compatibility of externalism and self-knowledge via introspection is blocked. Such a conclusion, to repeat, hardly rises to the level of definitive proof of compatibility. Substantiation of this considerably

stronger claim would require a program of empirical research that far exceeds the scope of this essay. Yet, it nonetheless should now be clear that in terms of theory there is no reason to accept that focusing upon the prevalence of slow switching cases substantively strengthens Boghossian's position vis-à-vis Warfield's objection. For while Ludlow's employment of premise (P3') does make Boghossian's argument valid, his modified version of the original argument has revealed itself to be suspect. This being so, IT remains unsubstantiated.<sup>16</sup>

#### REFERENCES

- Armstrong, J. 1994: 'Nations Before Nationalism', in *Nationalism*, ed. Hutchinson, J., and A. D. Smith, New York: Oxford University Press, 140-147.
- Boghossian, P. 1989: 'Content and Self-Knowledge', *Philosophical Topics* 17, 5-26.
- Burge, T. 1979: 'Externalism and the Mental', *Midwest Studies in Philosophy* 4, 73-122.
- Burge, T. 1988: 'Individualism and Self-Knowledge', *The Journal of Philosophy* 85, 649-663.
- Burge, T. 1998: 'Memory and Self-Knowledge', in *Externalism and Self-Knowledge*, ed. Ludlow, P., and N. Martin, Stanford, CA: CSLI Publications, 351-370.
- Davis, S. 2000: 'Social Externalism and Deference', in *Metarepresentations: A Multidisciplinary Perspective*, ed. D. Sperber, Oxford, Oxford University Press, 361-388.
- Dretske, F. 1970: 'Epistemic Operators', *The Journal of Philosophy* 67, 1007-1023.
- Ludlow, P. 1995: 'Externalism, Self-Knowledge, and the Prevalence of Slow Switching', *Analysis* 55, 45-49.
- Ludlow, P. 1997: 'On the Relevance of Slow Switching', *Analysis* 57, 285-286.

- Ludlow, P., and N. Martin. 1998: *Externalism and Self-Knowledge*, Stanford, CA: CSLI Publications.
- Nuccetelli, S. 2003: *New Essays on Semantic Externalism and Self-Knowledge*, Cambridge, MA, The MIT Press.
- Pappas, G. 2000: 'Epistemic Deference', *Acta Analytica* 15, 113-126.
- Pinker, S. 1994: *The Language Instinct*, New York, Penguin.
- Putnam, H. 1975: 'The Meaning of "Meaning"', in *Mind, Language and Reality*, New York, Cambridge University Press, 215-271.
- Sawyer, S. 1999: 'An Externalist Account of Introspective Knowledge,' *Pacific Philosophical Quarterly* 80, 358-378.
- Warfield, T. A. 1992: 'Privileged Self-Knowledge and Externalism are Compatible', *Analysis* 52, 232-237.
- Warfield, T. A. 1997: 'Externalism, Privileged Self-Knowledge, and the Irrelevance of Slow Switching', *Analysis* 57, 282-284.
- Wright, C., B. C. Smith, and C. MacDonald. 1998: *Knowing Our Own Minds*, New York, Oxford University Press.

## NOTES

<sup>1</sup> Three recent anthologies contain essays devoted precisely to this issue: Ludlow and Martin (1998); Wright, Smith, and MacDonald (1998); and Nuccetelli (2003).

<sup>2</sup> Warfield (1997) has already replied to Ludlow's modification of Boghossian's argument, albeit on different grounds than do I. According to Warfield, Ludlow does no more than Boghossian to show, beyond externalism being merely consistent with a lack of self-

---

knowledge, that the externalist doctrine actually *implies* as much. Warfield insists that to prove this stronger assertion Ludlow would have to show that incompatibility occurs not just in the actual world but in *every possible world*. Yet, it seems to me that the distinction between Warfield's response and Ludlow's argument comes down to this: Warfield has logical incompatibility in mind, while Ludlow has something more pedestrian in mind—i.e., that externalism and authoritative self-knowledge are incompatible in the sense that one cannot believe both. Ludlow (1997) does nonetheless update his version of IT in response to Warfield to state that what he specifically seeks to show is that 'externalism, authoritative self-knowledge, and slow-switching are mutually incompatible' (286). As we shall see, Ludlow's argument cannot hold up in either case: whether we are to take Ludlow as providing a more pedestrian account of incompatibility or a more specific account of this sort of mutual incompatibility.

<sup>3</sup> The argument I here attribute to Boghossian is actually a conflation of two distinct arguments within 'Content and Self-Knowledge'. The first (pp. 11-14) is meant to prove that externalism is incompatible with an agent knowing the contents of his/her thoughts via introspection, while the second (pp. 20-23) is meant to rule out the possibility that externalism is compatible with an agent knowing the contents of his/her thoughts based upon the fact that these thoughts are self-verifying. By explicating Boghossian's position in the manner that I have, I hope to provide what I take to be the sense in which these two arguments work together to purportedly accomplish the task of proving the incompatibility of externalism and authoritative self-knowledge in general.

<sup>4</sup> Boghossian does not exclude the possibility that one can end up with *both* Earthian and Twin-Earthian concepts. This option, as he puts it, 'is perfectly coherent—and a lot more

---

interesting...’ (13). However, he chooses to assume that slow switching causes wholesale displacement for two reasons: (1) this is the traditional way in which the story is told, hence it allows him to keep his argument as simple as possible; and (2) telling the story in the non-traditional way does not serve to circumvent his objections anyhow (cf. note 11, p. 24).

<sup>5</sup> The term ‘relevant alternative’ used here by Warfield was coined by Fred Dretske (1970), for whom ‘[a] relevant alternative is an alternative that might have been realized in the existing circumstances if the actual state of affairs had not materialized’ (1021). It is thus a possible state of affairs B that must be ruled out for one to be warranted in asserting that he/she possesses knowledge of state of affairs A. In other words, an alternative being merely logically possible does not in and of itself make it an alternative that can undermine our knowledge. Dretske proposes, for example, that there is little reason to believe that zebras at the zoo are actually mules that have been painted with stripes by the zoo authorities. Given that we have never run across a case in which this has occurred and can think of no reason why this would be done (‘Why should the zoo authorities do that?’ [1016]), such an alternative should not lead us to doubt our knowledge concerning whether it is the case that the creatures before us are indeed zebras. Not only is the painted mule alternative highly implausible, it is not relevant to an adequate explanation of what this particular creature might be. This is the sense in which I will use the term ‘relevant alternative’ within this essay.

<sup>6</sup> For a similar argument made on reliabilist grounds, see Sawyer (1999)—in particular, sections 6-7.

---

<sup>7</sup> It should be noted that routine and non-routine switches do not differ in degree of potency. Both the Biff case and the Twin Earth scenario have the same consequences if undetected. The difference between routine and non-routine switching consists instead in the fact that the former occurs far more frequently and regularly than the latter: only the former is thereby prevalent. Yet, it might be asked, why is this distinction of interest to the externalist? As I see it, it is a legitimate move for Ludlow to reframe the actuality of slow switching in terms of prevalence. My routine/non-routine switch distinction is simply a natural extension of this move.

<sup>8</sup> The externalist need not be concerned with questions concerning the *acquisition* of self-knowledge. In dealing with the issue of adaptation, and slow-switching cases more generally, that with which we are concerned is how such knowledge is *retained*. Notwithstanding the manner in which knowledge is acquired, the argument I here present thereby is intended merely to highlight that the *retention* of self-knowledge is not jeopardized by shifting linguistic communities.

<sup>9</sup> Davis provides a compelling argument that, given the truth of semantic externalism, the maintenance of a conditional disposition to defer is a necessary condition for concept possession. Insofar as this is so, as will become important later, it must likewise be a necessary condition for shifts in the content of such concepts.

<sup>10</sup> As Burge (1979) remarks, 'People are frequently held, and hold themselves, to the standards of their community when misuse or misunderstanding is at issue' (90). George Pappas (2000) provides perhaps the most direct substantiation of this point (cf. pp. 114-115) in a focused analysis of epistemic deference.

---

<sup>11</sup> This issue naturally raises yet more interesting questions that I unfortunately am not in a position to address within the confines of this essay: What of those occasions on which Biff is not aware that he has partial knowledge and thus does not defer? Are such occurrences possible (i.e., what is the basis for judgments of partial as opposed to authoritative knowledge)? Is it possible for him to be aware that he has partial knowledge yet to refuse to defer? Ludlow does allude to members within various communities—those who apparently act in a manner akin to what Stephen Pinker (1994) refers to as ‘language mavens’ (373)—being unwilling to defer to others (48).

<sup>12</sup> The plausibility of premise (P4) becomes more visible with Ludlow’s assertion that distinctive linguistic communities can be highly localized. ‘It is worth noting that the social groups that we defer to need not be sprawling networks of individuals, but may in fact be very local in scope’ (48).

<sup>13</sup> Armstrong discusses linguistic border guards within the context of a wider discussion of the development of nationalistic tendencies based upon ethnic delineations. Linguistic border guards often mark the geographical and hegemonic boundary of an ethnic group’s territory. They come in the form not only of different usages for the same term being employed by different groups but also of the use of wholly different words to refer to common concepts or objects. Although our discussion is considerably more general in form than is Armstrong’s, I here assume that his terminology is flexible enough to serve our purposes—if for no other reason than to at least provide a particularly lucid label to an otherwise abstract set of ideas.

<sup>14</sup> One may also question whether the dependence of my argument upon awareness of switches may itself beg the question to the extent that the Twin Earth scenario hinges

---

precisely upon lacking this sort of awareness. Given my specific concern with the issue of prevalent slow switching as describes by Ludlow, I do not believe my argument succumbs to such an objection. The suggestion of prevalence is, to my mind, what here alters the situation—and demands that I respond to Ludlow as I have.

<sup>15</sup> See Burge (1988) for a fuller discussion of our capacity to maintain this sort of self-attributional knowledge while nonetheless assuming the truth of externalism.

<sup>16</sup> An earlier draft of this essay was presented at the First Annual University of Rochester Graduate Conference in Epistemology, November 2001. I would like to thank attendees, in particular my commentator E.J. Coffman, for their excellent feedback. I am also very grateful for the insights and helpful criticisms offered by Peter Ludlow, Patrick Grim, and an anonymous reviewer.