

Jennifer Wilson Mulnix

University of Massachusetts Dartmouth, Department of Philosophy, 285 Old Westport Road,
US-02747 North Dartmouth, MA
jmulnix@umassd.edu.

**Semantic Externalism, Self-Knowledge, and
Slow Switching**

Abstract

Semantic externalism holds that the content of at least some of our thoughts is partly constituted by external factors. Accordingly, it leads to the unintuitive consequence that we must then often be mistaken in what we are thinking, and any kind of claim of privileged access must be given up. Those who deny that semantic externalists can retain any account of self-knowledge are ‘incompatibilists’, while those who defend the compatibility of self-knowledge with semantic externalism are ‘compatibilists’. This paper examines the claim of compatibilism, focusing on Burge’s “Slow Switching Argument” and Boghossian’s “Objection of Relevant Alternatives”. I argue that compatibilism is false, and that semantic externalism is incompatible with self-knowledge.

Key words

semantic externalism, anti-individualism, compatibilism, incompatibilism, slow switching, self-knowledge, epistemology, philosophy of mind.

Introduction

It is relatively uncontroversial that we have a kind of privileged access to our own mental states that we do not have to the mental states of others.¹ An asymmetry exists between first-person authority and third-person authority. Concerning this first-person privileged access, when we think of a tree, we may not know if there is an actual tree corresponding to the thought, but we do know that we are thinking of a tree. However, according to ‘semantic externalism’, meanings are not just in the head; instead, external objects are partly *constitutive* of thoughts.² My thought of a tree is only a thought of a tree if it is appropriately related (usually causally) to an actual tree. Defenders of semantic externalism are then charged with the unintuitive consequence that we must often be mistaken in what we are thinking. Furthermore, any kind of claim to privileged access must be given up. For, if mental states are constituted in part by the external environment, so that if the environment changes so does my thought, then how could I possibly have any claim to self-knowledge? Those who deny that semantic externalists can retain any account of self-knowledge are ‘incompatibilists’, while those who defend the compatibility of these two theses are known as ‘compatibilists’.

¹ Gilbert Ryle (*The Concept of Mind*, Hutchinson and Co, London 1949) is one famous exception.

² Though, as noted later, there are also forms of social externalism.

In this paper, I will examine the issue of compatibilism and whether semantic externalism can provide a plausible account of self-knowledge. In doing so, I will focus mainly on Tyler Burge's 'Slow Switching Argument' and the *Objection of Relevant Alternatives* offered by Paul Boghossian. I will argue that semantic externalism is incompatible with most robust accounts of self-knowledge, but that it *can* be made compatible with either a weak defeasible account or deflationary account. However, I also offer reasons to reject either of these accounts over the more standard models.

1. Self-Knowledge and Semantic Externalism

Before turning to the compatibilist issue, some prefatory remarks are in order. First, it is important to call attention to the central claims associated with the concept of 'self-knowledge'. Unfortunately, however, there exists significant disagreement on this score. Some hold that self-knowledge bears a strong connection with *a priori* knowledge,³ while others connect it with non-inferential knowledge.⁴ In most cases, self-knowledge is typically associated with some form of introspection.⁵ Moreover, self-knowledge is taken as a paradigmatic case of a second-order intentional state. In light of the variations among competing accounts of self-knowledge, I think we can, nonetheless, point to the following commitments of *self-knowledge*:

- (1) Self-knowledge is a different psychological state than the first-order state that is either the "object" or "content" of the self-knowledge.⁶
- (2) An internal component (narrow psychological state) is necessarily involved.
- (3) Either self-knowledge is not based on observation or it is not based on inference, or both.⁷
- (4) A subject has self-knowledge about at least some of her thoughts.
- (5) There is some kind of presumption in favor of a subject's claim to self-knowledge, though it need not be infallible.
- (6) The justification for self-knowledge is attained either through the subject's internal state or access, or through introspection being a reliable process.⁸

Elaborating on this final commitment, epistemic justification for self-knowledge can be analyzed on either an 'internalist' or 'externalist' account. On epistemic internalism, justification is sourced "inside" the agent. Setting aside inferential internalisms, the most common non-inferential internalisms will involve some kind of access or internal state that is the justification. Under an internal state account, a person would have knowledge of her mental states when they exemplify a particular non-relational property.⁹ Under an access account, a person would have knowledge of her own mental states if she had or could have access to the fact that her second-order belief is justified.¹⁰ Although under such internalist accounts it is unclear what internal state is required or what accessing justification would amount to, it is clear that justification is based on things internal to the subject. On epistemic externalism, on the other hand, non-inferential self-knowledge results from a reliable introspective belief-forming process.¹¹ Here, access on the part of a subject to her internal states is unnecessary for her to possess justification for her belief.

With this in mind, we are now ready to spell out the central commitments of *semantic externalism*:

3

Sarah Sawyer argues for a tight connection between self-knowledge and *a priori* knowledge, for instance (Sawyer, Sarah, “An Externalist Account of Introspective Knowledge”, *Pacific Philosophical Quarterly* 80 (1999), pp. 358–378). However, Tyler Burge draws a clear distinction between self-knowledge and *a priori* knowledge: “As one thinks a thought reflexively, it is an object of reference and knowledge, but simultaneously a constituent of one’s point of view. The essential role that the first-person singular plays in the epistemic status of authoritative *a priori* self-knowledge differentiates this knowledge not only from empirical knowledge, but also from most knowledge, the justification of which does not depend on the first-person point of view in the same way. The tendency to blur distinctions between *a priori* knowledge (or equally, knowledge involved in explication of one’s concept) and authoritative self-knowledge is, I think, an instance of Descartes’ central mistake: Exaggerating the implications of authoritative self-knowledge for impersonal knowledge of necessary truths” (Burge, Tyler, “Individualism and Self-Knowledge”, *The Journal of Philosophy* 85 (1988), pp. 649–663, here p. 662).

4

Paul Boghossian argues that self-knowledge is direct and without inference (Boghossian, Paul A., “Content and Self-Knowledge”, *Philosophical Topics* 17 (1989), pp. 5–26). Yet, James Chase argues that *some* self-knowledge is *inferential* (Chase, James, “Is Externalism about Content Inconsistent with Internalism about Justification?”, *Australasian Journal of Philosophy* 79 (2001), pp. 227–246).

5

Sarah Sawyer argues that introspection is “the means by which a subject has privileged access to her thoughts,” and so, introspection is intricately linked to self-knowledge (S. Sawyer, “An Externalist Account of Introspective Knowledge”, p. 358). But, Charles Landesman, Jr. argues there is a difference between introspection and self-knowledge, where introspection is the “act of observing one’s mental states” while self-knowledge is “the achievement of that knowledge.” He explains: “Few writings make the further distinction between the act of observing or attending to one’s present states for the purpose of acquiring knowledge of them and the actual achieving of or coming to possess such knowledge” (Landesman, Charles, Jr, “Consciousness”, in: Paul Edwards (ed.), *The Encyclopedia of Philosophy, Volume Two*, Macmillan, New York 1967, pp. 191–195). According to Vrinda Dalmiya, introspec-

tion does not entail a self-ascription of truths about one’s mind, but rather, implies a simple knowledge of one’s own mental occurrences, whereby a subject has privileged access to her thoughts: “Introspection, therefore, is a mental occurrence, which has as its object some other psychological state like perceiving, desiring, willing, feeling, etc. In being a distinct awareness-episode it is different from a more general ‘self-consciousness’ which characterizes all or some of our mental history” (Dalmiya, Vrinda, “Introspection”, in Jonathan Dancy & Ernest Sosa (eds.) *A Companion to Epistemology*, Blackwell, Oxford 1992, p. 218).

6

Not everyone agrees that the first-order thought is the “object” of the second-order thought, but will agree that the first-order thought provides the “content” for self-knowledge. In other words, “the thought that grass is green or grass is blue contains the content that grass is blue, but not the thought that grass is blue” (S. Sawyer, “An Externalist Account of Introspective Knowledge”, p. 377, n. 27).

7

This claim accommodates the fact that not all of our self-knowledge is *a priori*, and not all of it is non-inferential, but the account of self-knowledge we are interested in here will be either *a priori* or non-inferential. There are cases of which self-knowledge may be neither *a priori* nor non-inferential, but these cases can lay no claim to privileged access and thus, will not be a concern of this particular paper. I am only concerned with cases of self-knowledge that involve some sort of privileged access. Privileged access just is to be known either non-inferentially, non-experientially, or both.

8

One could substitute for “reliable process” some alternative externalist desideratum, such as “intellectual virtue”.

9

For example, when one’s psychological states are ‘self-presenting’, ‘self-evident’, ‘indubitable’, or ‘incorrigible’, such that one’s experience directly justifies the belief that one is having that experience.

10

Fumerton, Richard, *Metaepistemology and Skepticism*, Rowman and Littlefield, London 1995, especially Ch. 3.

11

I am assuming here that introspection (or whatever process is involved in self-knowl-



- (1) Intentional states are *constituted* in part by external factors (be they social or causal).
- (2) A significant variation in the environment of the speaker will necessitate a significant variation in the semantic content of the intentional state.
- (3) Two subjects may be in the exact same internal physiological state and be attributed two different psychological states. That is, the content of a significant range of psychological states does not supervene on the intrinsic features of their bearers.
- (4) The concepts and thoughts a subject can have depend essentially on the kinds of things in her environment. In other words, *meaning* depends in part on facts external to the subject.

One can then use this semantic externalist characterization of belief-states or intentional states to articulate the nature of both first-order and second-order belief-states. Specifically, first-order belief is constituted by three components: the object, its social or causal relation to the agent, and the agent's internal state.¹² The second-order belief, then, would be a combination of (1) the first-order belief (which is itself a combination of the object, the relation, and the internal state), (2) the relation between the first-order belief and the internal state, and (3) the internal state of the agent (which itself may not have intentional content). These three things constitute the second-order intentional state.

To illustrate the implications of semantic externalism, let us consider the following two scenarios, loosely derived from Hilary Putnam. Both scenarios involve two subjects named Bill and Phil, who both have a (first-order) belief about water and a (second-order) introspective belief about their first-order belief about water. In Scenario A, Bill is on Earth and Phil is on Twin Earth. On Earth, water is H₂O, and on Twin Earth water is XYZ. Only Bill may be correctly attributed the belief that he is thinking of water because 'water' means having a chemical composition of H₂O, among other things (and Phil may be attributed a thought of 'twin-water'). It is further stipulated that Bill and Phil have the same internal happenings going-on. However, even though Bill and Phil are in the same internal narrow psychological state, Bill and Phil have different beliefs. Furthermore, Bill and Phil both have a belief about *their belief about water*. Through introspection (or fill in your preferred counterpart of self-knowledge here), Bill and Phil both believe that they have a thought about water. The internal narrow psychological state of this second-order belief is also the same for Bill and Phil, as well as whatever *relation* is involved in the process of introspection (the relation between the first-order belief and the second-order internal state); however, in spite of this, their second-order beliefs are different due to varying first-order beliefs. Thus, two people could be in the same second-order internal state and yet have two different second-order introspective beliefs. For, Bill has a belief about water and Phil has a belief about twin-water. When Bill and Phil are reflecting on their beliefs about water, Bill is reflecting on his 'belief about water' while Phil is actually reflecting on his 'belief about twin-water'.

In Scenario B, on the other hand, Bill and Phil are both inhabitants of Earth. Furthermore, both were in contact with H₂O when forming their beliefs about water. In this scenario, their internal states are once again the same; however, this time the object of the first-order belief is also the same, and so, both Bill and Phil are thinking about water. Now, Bill and Phil are also thinking about their belief about water through introspection. Again, the second-order in-

ternal states of the two are qualitatively identical, but because the first-order belief is the same, the content of their second-order beliefs is also the same.

The only difference between the Scenario A and Scenario B is in the original object of the first-order thought. In both scenarios, the internal states on the first and second levels were the same. Everything “internal” to Bill and Phil was qualitatively identical in Scenarios A and B. And yet, in Scenario A, the second-order beliefs of Bill and Phil varied, while in Scenario B they were the same. And the only explanation for this is the *actual object* of the first-order belief (or in Burge’s case, the social community’s understanding of ‘water’). Yet, the semantic externalist wants to claim that Bill and Phil have direct access to their first-order beliefs *independent* of any awareness of their specific *environment*. Moreover, we have no indication of any change among the internal narrow psychological component in either scenario. And in neither scenario is Bill or Phil able to determine the content of his belief by the quality or nature of his internal state.

Immediately, we can see how epistemic internalism poses a real threat to compatibilism: according to epistemic internalism, for self-knowledge of mental states to be possible, the internal state must in some sense “disqualify” false beliefs. That is, the internal state that justifies Bill in believing he has a thought of water cannot be the same internal state that would accompany Bill’s *false* belief that he has a thought of twin-water. Clearly, this renders the alleged compatibility claim between self-knowledge and semantic externalism highly unlikely on any internalist understanding of self-knowledge.¹³ Under an externalist account of self-knowledge, we would ask whether the introspective process that outputs Bill and Phil’s second-order belief is a reliable faculty in these circumstances. It seems that, as long as Bill and Phil acquire the beliefs in the appropriate way (reliably), then their second-order beliefs

edge) is a non-inferential belief-forming process, though this remains an open question. It is unclear whether the inputs to one’s introspective process are beliefs or non-doxastic internal states. Nonetheless, even if the inputs to one’s introspective process are (first-order) beliefs, this does not automatically render the outputted second-order introspected beliefs as inferential in character.

12

By ‘belief-state’ I shall mean an intentional state with intentional content. Other intentional states such as desires, hopes, etc. are also constituted by the three aforementioned components. Any of these states, including belief-states, are ‘psychological states’. Here, psychological-state is an intentional state and is not to be confused with Putnam’s ‘narrow psychological states’. Putnam’s narrow psychological states are actually what I have called the agent’s ‘internal states’ and are not proper psychological states. Internal states may be images, goings-on, or other things internal to the agent that do not have intentional content. Externalists claim that there is nothing about the internal state that makes it a representation of anything, even if the internal states include resembling images. Thus, externalists do not deny that there are “things” that go on in the head of a person,

but they deny that this internal state alone is sufficient for representation.

13

However, James Chase argues that semantic externalism is compatible with internalist accounts of self-knowledge, as long as some self-knowledge is only “narrow” in character and does not involve any wide concepts. If this were the case, then all factors would be internally available to an agent because the contents involved would all be narrowly individuated. Essentially, Chase is drawing attention to the fact that not all semantic externalists need accept the anti-individuation thesis regarding mental content, or at least, that some semantic externalists can maintain that some content (the mental content of self-knowledge) is narrowly individuated even though meaning is in some sense external. And of course, the consistency Chase attributes between semantic externalism and epistemic internalism would also hold true between semantic externalism and epistemic externalism. Hence, if we accept the claim that self-knowledge involves only narrow content, then we have a *prima facie* reason to embrace compatibilism, even for an epistemic internalist (J. Chase, “Is Externalism about Content Inconsistent with Internalism about Justification?”, pp. 227–246).

are justified, regardless of whether that belief is of water or twin-water. But, as will be clear shortly, one must also ask whether the presence of twin-water in one environment creates a ‘relevant alternative’ that significantly weakens Bill or Phil’s initial justification. I will attempt to answer this further question in connection with Boghossian’s *Objection of Relevant Alternatives*. But first, I will articulate Burge’s defense of compatibilism in the context of his ‘Slow Switching Argument’.¹⁴

2. Compatibilism and the Slow Switching Argument

Turning now directly to the issue of compatibilism, we are interested in answering the question of whether semantic externalism precludes the ability for one to have self-knowledge. Under semantic externalism, is there the same kind of room for error in an agent’s second-order introspective belief that occurs at the first-order belief level? That is, semantic externalism holds that a variation in the object for the first-order belief-state will change the semantic content of the belief-state, all of which may be unknown to the subject; thus, can there be a variation of first-order belief-states that change the semantic content of the second-order belief states of which the subject would *also* be unaware? In other words, must the contents of the two states always match? This seems to be the question the semantic externalist is charged with answering.

Burge answers this question by asking us to consider the following scenario, according to which there is only Bill, but where Bill is now traveling between Earth and Twin Earth. ‘Slow Switching’, according to Burge, occurs when, in the process of vacillating between home and twin environments, one stays long enough in one situation for the thoughts to change. In regular or quick switching, on the other hand, not enough time is spent in either environment to alter the subject’s thoughts. Furthermore, Burge imagines a case of Slow Switching where the agent is unaware of the switching going on between worlds – he does not ever know which of the two worlds he is in (and the continuity of his life is not disrupted) – and the thoughts change due to acquired concepts and perceptions, but where this is unknown to the subject:

“[I]magine a case of slow switching between actual home and actual twin-home situations. In the former situation, the person may think “I am thinking that water is a liquid.” In the latter situation, the person may think “I am thinking that [twin-]water is a liquid.” In both cases, the person is right and fully justified as ever. *The fact that the person does not know that a switch has occurred is irrelevant to the truth and justified character of these judgments.* Of course, the person may learn about the switches and ask “Was I thinking yesterday about water or [twin-]water?” – and not know the answer. Here knowing the answer may sometimes indeed depend on knowing empirical background conditions. ... the answer is obviously “both.” Both concepts are used. Given that the thought is fixed and that the person is thinking it self-consciously, no new knowledge about the thought could undermine the self-ascription – or therefore its justification or authority.”¹⁵

Moreover, the internal narrow psychological state is invariant – the inner goings-on remain the same and everything “feels” the same. “The root idea is that at least some aspects of one’s mental life are fixed by the chemical composition of one’s body. One might call these aspect *pure phenomenological feels*.”¹⁶ But, according to Burge, even though the person would not be able to tell you when slow switching has occurred – he cannot tell the actual environment from various twin environments – this does not mean he does not know what thoughts he has. Bill can have direct, authoritative knowledge of his thoughts without being required (or able) to know his external environment. Yet, it is hard to see how this can be the case. So, why does Burge maintain this?

It might help here to ask the following question: When Bill is reflecting on his belief that there is water, what is the “object” of his second-order belief? Is it the first-order belief (‘I believe there is water’); the content of the first-order belief (‘there is water’); or the internal state that is a partial constituent of the first-order belief? Burge’s answer is that it is a mistake to think of self-knowledge as simply taking some kind of “object” before the mind.¹⁷ Instead, self-knowledge consists in thinking one’s (first-order) thought self-ascriptively: “One knows one’s thought to be what it is simply by thinking it while exercising second-order, self-ascriptive powers.”¹⁸ Burge explains: “When one knows that one is thinking that *p*, one is not taking one’s thought (or thinking) that *p* merely as an object. One is thinking that *p* in the very event of thinking knowledgeably that one is thinking it. It is thought and thought about *in the same mental act*. So any conditions that are necessary to thinking that *p* will be equally necessary to the relevant knowledge that one is thinking that *p*.”¹⁹ Thus, “such knowledge [of one’s own mental events] consists in a reflexive judgment which involves thinking a first-order thought that the judgment itself is about. The reflexive judgment simply inherits the content of the first-order thought.”²⁰ Burge elaborates:

“Basic self-knowledge is self-referential in a way that insures that the object of reference just is the thought being thought. If background conditions are different enough so that there is another object of reference in one’s self-referential thinking, they are also different enough so that there is another thought. The person would remain in the same reflexive position with respect to his thoughts and would again know, in the authoritative way, what he is thinking.”²¹

Thus, “one has no ‘criterion’, or test, or procedure for identifying the thought, and one need not exercise comparisons between it and other thoughts in order

14

It should be noted that not all semantic externalists find Burge’s ‘Slow Switching Argument’ compelling as a defense of semantic externalism. For instance, Sarah Sawyer claims that the ‘Slow Switching Argument’ is entirely implausible: “The proposal is plausible only if it is assumed that [Bill] would still be employing natural kind terms and concepts even though [his] environment contained multiple duplicates. This assumption could be contested on the grounds that the ability to acquire and deploy natural kind terms and concepts may presuppose that the natural kind substances in our environments can be distinguished by and large by their manifest properties. That is, it may be that were duplicates commonplace, our terms and concepts would be purely descriptive, and would not relate to specific natural kinds at all” (S. Sawyer, “An Externalist Account of Introspective Knowledge”, p. 372). Moreover, she argues: “It is often likewise assumed that being switched between duplicate environments on a frequent basis would not disrupt [Bill’s] psychological functions. This is far from obvious. The proper functioning of a mind may require a broadly stable environment. The root of the claim that the mind should be seen as extended onto the world is that the mind is not essentially detachable from the world. This seems to have been overlooked by those who make frequent appeal to so-called switching cases” (Ibid., p. 377, n. 20).

15

T. Burge, “Individualism and Self-Knowledge”, p. 659, italics added.

16

Ibid., p. 653.

17

Burge argues that thinking of one’s thoughts as “objects” (either like physical objects or as objects of thought *qua* propositions) before the mind is deeply misconceived (Ibid., p. 660). Donald Davidson, agrees: “[W]e should deny that there are objects of the mind. The source of the trouble is the dogma that to have a thought is to have an object before the mind” (Davidson, Donald, “Knowing One’s Own Mind”, *Proceedings and Addresses of the American Philosophical Association* 60 (1987), pp. 441–458, p. 455).

18

T. Burge, “Individualism and Self-Knowledge,” p. 656.

19

Ibid., p. 654.

20

Ibid., p. 656.

21

Ibid., p. 659.

to know it as the thought one is thinking. Getting the ‘right’ one is simply a matter of thinking the thought in the relevant reflexive way.”²²

According to Burge, the truth-maker for self-knowledge is part of the very belief itself – the second-order belief that constitutes self-knowledge contains its own truth-maker for its content. Because the introspective belief is reflexive and self-referential, its *content* is “logically” locked by the first-order belief. That is, the content of the second-order thought is locked to the content of the first-order thought, even though the first-order thought is not the *object* of the second-order thought. Burge explains:

“In basic self-knowledge, one simultaneously thinks through a first-order thought (that water is a liquid) and thinks about it as one’s own. The content of the first-order (contained) thought is fixed by non-individualistic background conditions. And by its reflexive, self-referential character, the content of the second-order judgment is logically locked (self-referentially) onto the first-order content which it both contains and takes as its subject matter. Since counterfeit contents logically cannot undermine such self-knowledge, there should be no temptation to think that, in order to have such knowledge, one needs to master its enabling conditions.”²³

Burge’s thesis, then, is a claim about the matching of meaning or content. Accordingly, if the object (of the first-order belief) has changed, then the first-order belief has also undergone a change, and by necessity, the second-order belief must change as well. We see, then, that Burge’s conclusion relies on the claim that there is a sort of *logical* connection between the content of our first-order beliefs and second-order beliefs. According to Burge, second-order beliefs, by their natures, are self-referential, and thus, the content of the first-order belief is *necessarily* fixed in the second-order belief.

These are very strong claims made by Burge, and they must be examined carefully. For it is clear that Burge is not simply saying it is *possible* to have self-knowledge, but rather, it is *logically guaranteed*. This is due to the fact that Burge does not consider introspection to be a distinct act. Burge maintains that self-knowledge is not a *relation* between two acts (the first-order belief and the second-order belief), but is, instead, in the second-order belief alone. This kind of account of self-knowledge is a *deflationary* account. A deflationary account of self-knowledge holds that self-knowledge is not a cognitive achievement.

“Ordinarily, to know some contingent proposition you need either to make some observation, or to perform some inference based on some observation. In this sense, we may say that ordinary empirical knowledge is always a *cognitive achievement* and its epistemology always *substantial*.”²⁴

Because Burge’s deflationary self-knowledge is less substantial, it is easier to claim compatibility with semantic externalism. Andre Gallois characterizes a deflationary account of self-knowledge as follows:

“They [Burge and Davidson] would concede that the following is an implication of externalism. In the case of many propositional attitudes, in one sense, we cannot know which ones we have without conducting an empirical investigation. However, that does not preclude our having first-person authority over those propositional attitudes. It does not, principally because one type of error is excluded when it comes to present-tense first-personal self ascriptions of propositional attitudes. A propositional attitude that one has cannot masquerade as having a content it does not possess. There is no room for facsimile-based error in the case of present-tense first-personal self ascriptions of propositional attitudes.”²⁵

Under Burge’s account, then, there is no room for a “mismatch” between the content of the first-order thought and the second-order thought, because only one mental act is involved. Thus, the only way for the content of the

second-order belief *not* to “track” the content of the first-order belief is if the statement “I believe that water is wet” is *false* on Twin Earth. In other words, had there been *no* water at all, my thought that ‘I believe that water is wet’ would be false. Hence, Burge claims that the compatibility between semantic externalism and self-knowledge is met by a *deflationary* account, according to which one is *ensured* first-person authority over her mental states.

An important consequence of Burge’s deflationary account, then, is that a subject can *never* be in error about what she thinks. However, denying the possibility of error outright seems highly unintuitive. Moreover, it would follow that thinking that one thinks that *p* entails thinking that *p*. And yet, this does not seem to be the case either. For instance, judgments of one’s standing mental states, or judgments about what one occurrently desires or fears, are not self-verifying. I may judge that I believe that writing requires concentration, and yet I need not actually believe that writing requires concentration in order to think this thought. Moreover, I may judge that I fear that writing requires concentration without actually fearing that it does.²⁶ Thus, Burge’s inability to allow for any mistaken belief ascriptions seems enough to render Burge’s account implausible. What is more, thinking that one thinks that *p* is *not* simply thinking that *p* in a special way. The argument behind this is simple: “If it were, whatever renders one’s thought that *p* false would render one’s thought that *one thinks* that *p* false. ... even if writing does not require concentration, I need not be mistaken to think that I think that it does.”²⁷ Thus, the burden of proof seems to be placed upon Burge, for the incompatibilist does not need to show that the content of a first-order thought could correspond to an entirely different content of the second-order thought, but Burge needs to show that it *cannot*.

Another key problem with deflationary accounts of self-knowledge is that there is the possibility that there is *no* first-order psychological state at all, or that it is of a different *type* than the one conceived.²⁸ And guaranteeing that the first-order belief exists and is of the correct type is equally as important as guaranteeing a match of content. McKinsey states, “there is no reason to suppose that the agent would have privileged access even to the fact that the episode is an episode of *thought*, as opposed to being, say, an episode of indigestion.”²⁹ Intuitively, however, it *does* seem that when we are reflecting on our own beliefs, about say, water, for example, we are in part reflecting on our *belief* in addition to its content. Therefore, match of content between first-order belief and second-order belief not sufficient to demonstrate authoritative self-knowledge.

22
Ibid., p. 656.

23
Ibid., pp. 659–660.

24
P. Boghossian, “Content and Self-Knowledge”, p. 17.

25
Gallois, Andre, *The World Without, The Mind Within: An Essay on First-Person Authority*, Cambridge UP, Cambridge 1996, p. 177.

26
P. Boghossian, “Content and Self-Knowledge”, p. 21.

27
A. Gallois, *The World Without, The Mind Within*, p. 173.

28
Ibid., p. 168. See also: S. Sawyer, “An Externalist Account of Introspective Knowledge”; and McKinsey, Michael, “Anti-Individualism and Privileged Access”, *Analysis* 51 (1991), pp. 9–16.

29
M. McKinsey, “Anti-Individualism and Privileged Access”, p. 11.

Hence, despite the possibility that a deflationary account of self-knowledge might be compatible with semantic externalism, there are compelling reasons for disregarding Burge's deflationary account as an account of genuine self-knowledge. More specifically, Burge's account requires infallibility, such that, *logically*, there is no room for error, and this makes the account too strong. What is more, Burge denies that self-knowledge is a cognitive achievement, and in this respect, his deflationary account is also too weak, since it cannot ensure knowledge of the *existence* or *type* of mental state, but only of its content. There is something highly unintuitive about an account of self-knowledge that renders it cognitively insubstantial. However, if we think that self-knowledge is a *distinct* psychological state from the original first-order psychological state, then Burge's view must respond to the objection that self-knowledge is not a logically infallible way of knowing. So, on these other non-deflationary accounts of self-knowledge – according to which self-knowledge is a distinct psychological state – does Bill know his second-order belief content when he is subjected to Slow Switching?

Well, one puzzle immediately emerges in connection with the *belief* component of self-knowledge. Namely, how is it that we are to suppose that Bill ever has a *belief* about twin-water? It seems that a person must first have the requisite concepts in order to have (second-order) self-knowledge of a first-order belief. Charles Landesman elaborates: “it is possible that the concepts used in the description of a mental state are not possessed by the individual who is in that state. In such a case the individual could not possess self-knowledge.”³⁰ Of course, the semantic externalist may respond that we are able to attribute to Bill a that-clause containing “twin-water” once Bill has been in a community which standardly uses the concept of water in this deviant way (or once Bill has been appropriately causally situated in a world with only XYZ and where the concept of water is used in this way). But, when Bill says, “I have a belief about water” does Bill have only one concept in his mind, or two? According to Burge, Bill has *two* concepts available to him under the Slow Switching thought experiment, even if Bill does not know this.

Yet, once again, the consequence of this position is that it may be the case that Bill is often mistaken about what he believes. If Bill is not an inhabitant of Twin Earth and has never previously encountered XYZ, then upon initial interaction with the substance, he would not have a belief about water even if he believed himself to have one. That is, if when looking at XYZ, Bill says “that is water,” he does not actually have a belief about water, but instead, about twin-water. Bill does not know that the substance is XYZ, nor does he know of the concept of twin-water.³¹ And, if Bill has no belief about twin-water, then he cannot be said to have self-knowledge of the sort “I have a belief about water” where ‘water’ is used on Twin Earth. Yet, the semantic externalist's fundamental thesis is that Bill *does* have a belief about twin-water, because all that this entails is that Bill is in the appropriate community (Twin Earth), and so, Bill will have the requisite concept. On Burge's account, then, the *belief* component for self-knowledge would be fulfilled. However, *even if* Bill is in possession of the requisite Twin Earth concept, it is still far from clear how Bill is to differentiate among his twin concepts in different environments. This seems to be the crux of the incompatibilist charge against the semantic externalist. To evaluate this charge, I will now turn to a consideration of Boghossian's *Objection of Relevant Alternatives*.

3. The Objection of Relevant Alternatives

The motivation behind Burge's 'Slow Switching Argument' is to make twin-water a relevant alternative to water, while also remaining (non-inferentially) non-discriminable to Bill. Recall that Bill is slowly switched back and forth between Earth and Twin Earth, though he is unaware that the switches are taking place (in fact, Bill may not even be aware that there is a Twin Earth). Bill is unable to distinguish water from twin-water by their superficial characteristics, and all his internal goings-on remain the same when on Earth and Twin Earth – it is only the underlying structure of the object that changes. Nonetheless, by remaining long enough in each environment, Bill has acquired both Earth and Twin Earth concepts, 'water' and 'twin-water' (though Bill is unable to distinguish the concept 'water' from 'twin-water'). And, once Bill has acquired both concepts, twin-water must be considered a relevant alternative. Burge wants to conclude from this that Bill is able to have self-knowledge of his thought of twin-water because he possesses the requisite concept. Thus, the idea is that, if twin-water is a relevant alternative, and yet, Bill can have first-person authority nonetheless, then compatibilism will have been established. Put another way, by allowing twin-water as a relevant alternative, Burge grants almost everything to the incompatibilist, while also claiming that the compatibility between self-knowledge and semantic externalism has been established.

It is important to note, however, that Burge must go farther than simply demonstrating that the twin concept was available. Burge must now be required to *exclude* one of twin concepts as a relevant alternative in his assessment of Bill's self-knowledge. But, this would require Bill to *reason* to a conclusion, which would require Bill to refer to his *environment*. However, since it has been presupposed that Bill may not even be aware of the existence of a Twin Earth or that any kind of switch is occurring, how can Bill utilize evidence from his environment? Or, is it the case that Bill *always* has *both* concepts at hand? Again, Burge claims that both concepts are used. Of course, if this were the case, then this creates a broader worry over what makes it true which concept would be applicable. According to Boghossian, there seems to be no easy answer: "It certainly does not seem right to say...that it is simply a function of the environment in which the thought is tokened. Nor are there obvious dimensions of difference to appeal to."³² Alternatively, we might think that only one concept is available at any given time because "with every such slow switch a wholesale displacement of [Bill's] resident concepts takes place."³³ If this were the case, then Bill does not know he has the concept 'water' (as opposed to 'twin-water') unless he knows he is in a world with water (and not twin-water). If Bill were to know he was in a world with water, then we may assume Bill could know he has the concept of water, though this knowledge would be *a posteriori* – this kind of knowledge can only be gained by examining his environment, and it would not be a case of self-knowledge.

³⁰ C. Landesman, "Consciousness", p. 192.

³¹ Of course, Bill would not have the word 'twin-water', but this is beside the issue. The word is used merely as a heuristic device to make the two concepts distinct from each other, since both would, in some sense, be contained in the word 'water'.

³² P. Boghossian, "Content and Self-Knowledge", p. 24, n. 11.

³³ *Ibid.*, p. 13.

Boghossian's *Objection of Relevant Alternatives* is a straightforward epistemological argument: in order to be justified in believing something, there must be no relevant alternatives available that would defeat the justification. Put another way, "a person is said to know that p just in case he *distinguishes* or *discriminates* that truth of p from relevant alternatives."³⁴ But what makes something a *relevant* alternative? Consider the example of Henry, who is driving along the countryside looking at barns. Does Henry *know* that he is looking at a particular barn? Probably so, if the countryside is filled with barns, and Henry is able to see one from a short distance in good lighting. But, if we now add to the story that the entire countryside is also replete with barn *facsimiles*, then it seems that the answer is now: No, Henry does *not* know that he sees a barn, even if the one he happens to be looking at is a real barn and not a facsimile. The reason for this is an addition of a relevant alternative, the *barn facsimile*. If one were to inform Henry of the presence of barn facsimiles, he may no longer be sure of his belief. Even aside from Henry's own assurance, once we are made aware of the presence of facsimiles, we would no longer attribute justification to Henry's belief. In sum, "what the presence of the facsimiles does is make this possibility *relevant*; or it makes us *consider* it relevant."³⁵

The key here is Henry's inability to *distinguish* the counterfactual state of affairs from the actual state of affairs. That is, "if there is a relevant possible state of affairs in which p is false and which is indistinguishable by him from the actual state of affairs, then he fails to know that p ."³⁶ If, in the original situation, there were no barn facsimiles but only real barns, then there would be no relevant possible state of affairs in which p is false and yet indistinguishable from the actual state of affairs. And, in a countryside with only barns, Henry would know that he sees a barn. Thus, Henry would be said to know p provided p was true, and "there is no relevant contrary q of p such that, if q were true (rather than p), then S would (still) believe that p ."³⁷

However, when does something become a relevant alternative and not just an idle one? It may be that there is no set of relevant alternatives which is uniquely determined by the subject's circumstances. For instance, if there were no barn facsimiles in that particular countryside, but instead, in a countryside two miles away, would this still make barn facsimiles a relevant alternative? This vagueness could pose a problem for determining whether, in Burge's 'Slow Switching Argument', the twin concepts are relevant alternatives. Nonetheless, general concerns aside, it seems quite clear that in Burge's 'Slow Switching Argument' the concept twin-water *is* a relevant alternative (and not merely, an idle alternative) to Bill's thought of water. Ted Warfield disagrees. He defends the compatibility of semantic externalism with self-knowledge by arguing that in Burge's Slow Switching thought experiment the twin concepts are *not* relevant alternatives. Warfield asserts that even if Slow Switching cases exist, twin concepts are never relevant alternatives, and hence, semantic externalism is compatible with self-knowledge.

Warfield defends his claim by first formalizing Boghossian's argument against compatibilism 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 twin-water thoughts.
- (P3) If the Switching Case is actual, then twin-water thoughts are relevant alternatives of water thoughts.
- (C1) S doesn't know that P by introspection.³⁸

However, Warfield argues, C1 does not follow from the three premises. Instead, the only conclusion Boghossian can draw is C1*:

(C1*) If the Switching Case is actual then S doesn't know that P by introspection.

Warfield then claims that C1* has no bearing on the issue of compatibilism. The only relevance Boghossian's conclusion has is to the question: "Given externalism, is it necessary that the contents of a thinker's thoughts are knowable to the thinker on the basis of introspection?"³⁹ And, the answer of "No" does no harm to a compatibilist, Warfield argues. In other words, "this shows at most that externalism is *consistent with* a lack of self-knowledge; it does not show that externalism *implies* a lack of knowledge."⁴⁰

But, I believe Warfield's defense of compatibilism is hasty and incomplete. As Peter Ludlow argues, switching cases may be actual, and in fact, *commonplace*.⁴¹ Thus, Warfield cannot simply rule out twin concepts as relevant alternatives. Warfield's response is to claim that Ludlow's argument only demonstrates that in the *actual* world semantic externalism is incompatible with self-knowledge. But, "to show that these doctrines are incompatible one needs to show that every possible world in which externalism is true is a world in which individuals do not have privileged self-knowledge."⁴² By showing that in one world – the actual world – some people cannot have privileged self-knowledge because of switching cases does not demonstrate the truth

34

Goldman, Alvin I., "Discrimination and Perceptual Knowledge", *The Journal of Philosophy* 73 (1976), pp. 771–791, p. 772. It should be noted that this remark is made in the context of perceptual knowledge rather than self-knowledge, but this will make little difference for the purpose to which it is used here.

35

A. Goldman, "Discrimination and Perceptual Knowledge", p. 775. Nozick's account of justification may also prove to be of some help here. Briefly, Nozick's account of justification is as follows: (1) *p* is true; (2) S believes, via method or way of coming to believe M, that *p*; (3) If *p* weren't true and S were to use M to arrive at a belief whether (or not) *p*, then S wouldn't believe, via M, that *p*; and (4) If *p* were true and S were to use M to arrive at a belief whether (or not) *p*, then S would believe, via M, that *p*. So, "S knows that *p* if and only if there is a method M such that (a) he knows that *p* via M, his belief via M that *p* satisfies conditions 1-4, and (b) all other methods M1 via which he believes that *p* that do not satisfy conditions 1-4 are outweighed by M" (Nozick, Robert, *Philosophical Explanations*, Harvard UP, Cambridge 1981, p. 179). Understood in this way, it would also be the case that Henry would not know that he sees a real barn if the countryside was replete with barn facsimiles.

36

A. Goldman, "Discrimination and Perceptual Knowledge", p. 774.

37

Ibid., p. 778.

38

Warfield, Ted A., "Privileged Self-Knowledge and Externalism are Compatible", *Analysis* 52 (1992), pp. 232–237, here pp. 234–235.

39

Ibid., p. 235.

40

Warfield, Ted A., "Externalism, Self-Knowledge, and the Irrelevance of Slow Switching", *Analysis* 57 (1997), pp. 282–284, here pp. 282–283.

41

Ludlow, Peter, "Externalism, Self-Knowledge and the Prevalence of Slow Switching", *Analysis* 55 (1995), pp. 45–49. We might also worry about Warfield's claim for the reason that he draws a distinction between "*a priori* introspective (privileged) self-knowledge" and "(stronger) *a priori* knowledge that does not include introspection." Interestingly, it is the former kind of self-knowledge which interests Warfield, as he believes the latter kind of self-knowledge is *incompatible* with semantic externalism. In fact, Warfield concedes that if people do actually have the stronger *a priori* type of self-knowledge of their mental states, then compatibilism is in danger.

42

T. Warfield, "Externalism, Self-Knowledge, and the Irrelevance of Slow Switching", p. 284.

of incompatibilism, Warfield concludes. Ludlow points out, however, that if there is incompatibility in *any nearby* world, then compatibilism has to be surrendered.

Ludlow's response seems to me to be correct. Using a counterfactual analysis of justification, if twin concepts are relevant alternatives in very nearby worlds, this undermines the justification of beliefs in the actual world. Or rather, as Ludlow argues, if twin concepts are relevant in the actual world, then this undermines the justification for beliefs in any nearby possible world. Thus, no *convincing* argument has been given for why twin-water should not be considered a relevant alternative. And barring such argument, this would defeat the semantic externalist's claim to accommodate self-knowledge on any robust account. Of course, provided a *weak enough* account of defeasible justification, semantic externalists can argue that their view is compatible with self-knowledge on *either* epistemic externalism or certain epistemic internalisms.⁴³

Yet, what about when slow switching does *not* occur, so that Bill is only in the actual world, or the switching that takes place between worlds is too quick? It seems in these cases that twin-water would *not* be a relevant alternative, and it would be unnecessary for Bill to examine his environment in order to know that he is looking at water or that he has the concept of water. Thus, in these situations, there appears to be no incompatibility between semantic externalism and self-knowledge. But what about Bill's privileged access? In other words, how does this accommodate our commonsense intuition that Bill's knowledge is privileged or could be obtained solely through introspection?⁴⁴ What the semantic externalist then needs to demonstrate is how introspective evidence can be sufficient to justify claims of knowledge of thought content, even though introspection underdetermines thought content – introspection cannot by itself individuate thought contents. And while Bill may be able to discriminate his thought of water from his thought of gin and his thought of grass, for example, he will *not* be able to distinguish his thought of water from his thought of twin-water on introspective evidence. Hence, Bill's knowledge based on introspective evidence must be *defeasible*, since his evidence will be inconclusive when Bill is entertaining many duplicate thoughts.⁴⁵ And yet this is an unintuitive result: self-knowledge is typically associated with a privileged access to one's own thoughts, where there is a clear presumption in favor of a subject's claim to self-knowledge.

In sum, the account of self-knowledge most compatible with semantic externalism is a type of deflationary account of self-knowledge. Yet, a deflationary account was shown to be highly problematic as an account of self-knowledge: it is too strong, since *logically* there is no room for error; but it is also too weak, since it cannot ensure knowledge of the *existence* or *type* of mental state, but only of its content. Finally, there is something highly unintuitive about an account of self-knowledge that renders it cognitively insubstantial. When we considered non-deflationary accounts of self-knowledge, we faced the problem of relevant alternatives, which renders the justification as *defeasible* at best. However, provided a *weak enough* account of defeasible justification, semantic externalists can argue that their view is compatible with self-knowledge. But the price paid for this compatibility is an account of self-knowledge which is at odds with our commitment to privileged access and first-person authority. Thus, in order to render semantic externalism compatible with self-knowledge, our intuitive understanding of self-knowledge will have been corrupted.

Jennifer Wilson Mulnix

Semantički eksternalizam, samospoznaja i *slow switching*

Sažetak

Semantički eksternalizam tvrdi da je sadržaj barem jednog dijela naših misli dijelom uspostavljen na izvanjskim faktorima. Na temelju toga dolazi se do neintuitivne posljedice da naše misli često moraju biti zablude, i da moramo odbaciti bilo kakvu težnju za privilegiranim pristupom. Oni koji niječu da semantički eksternalisti mogu zadržati bilo kakav pojam samospoznaje su »inkompatibilisti«, a oni koji brane kompatibilnost samospoznaje sa semantičkim eksternalizmom nazivaju se »kompatibilisti«. Ovaj rad ispituje kompatibilističke tvrdnje fokusirajući se na Burgeov »slow-switching argument« i Boghossianov »prigovor relevantnih alternativa«. Tvrdim da je kompatibilizam pogrešan, tj. da je semantički eksternalizam nekompatibilan sa samospoznajom.

Ključne riječi

semantički eksternalizam, anti-individualizam, kompatibilizam, inkompatibilizam, *slow switching*, samospoznaja, epistemologija, filozofija uma

Jennifer Wilson Mulnix

Semantischer Externalismus, Selbsterkenntnis und *Slow Switching*

Zusammenfassung

Der semantische Externalismus behauptet, der Inhalt mindestens einiger unserer Gedanken sei partiell durch externe Faktoren konstituiert. Hiernach resultiert dies mit der unintuitiven Folge, dass wir uns des Öfteren in eigenen Überlegungen täuschen, und uns von jeglichem Anspruch auf einen privilegierten Zugang lossagen müssen. Jene, die bestreiten, die semantischen Externalisten könnten irgendein Bild von der Selbsterkenntnis behalten, sind „Inkompatibilisten“, während jene, die die Kompatibilität der Selbsterkenntnis mit dem semantischen Externalismus verteidigen, als „Kompatibilisten“ zu bezeichnen sind. Der Artikel prüft die Überzeugungen des Kompatibilismus, sich fokussierend auf Burges „Slow-Switching-Argument“ sowie Boghossians „Einwand der relevanten Alternativen“. Ich vertrete die Ansicht, der Kompatibilismus sei falsch, bzw. der semantische Externalismus sei mit der Selbsterkenntnis inkompatibel.

Schlüsselwörter

semantischer Externalismus, Antiindividualismus, Kompatibilismus, Inkompabilismus, *Slow Switching*, Selbsterkenntnis, Epistemologie, Philosophie des Geistes

43

For instance, on a weak access internalist model, when Bill introspects that he has a thought of water, his internal state would defeasibly justify his belief and render it a case of self-knowledge, for the reason that Bill need only have potential access (which is itself not constitutive of the justification) to the conditions that constitute his justification. However, under a stronger acquaintance internalist model, a subject's non-inferential justification "is constituted by a direct relation that obtains between a belief and the fact that makes true the belief," and so, this

stronger view would remain incompatible with semantic externalism (R. Fumerton, *Metaepistemology and Skepticism*, p. 185). The compatibility of semantic externalism with some forms of epistemic internalism, then, remains highly suspect.

44

S. Sawyer, "An Externalist Account of Introspective Knowledge", p. 369.

45

Ibid., p. 370.

Jennifer Wilson Mulnix

Externalisme sémantique, connaissance de soi et *slow switching*

Résumé

L'externalisme sémantique affirme que le contenu d'au moins quelques-unes de nos pensées est en partie constitué par des facteurs externes. Par conséquent, il mène à la conséquence non-intuitive que nous devons souvent nous tromper sur ce que nous pensons et que toute prétention à un accès privilégié doit être abandonnée. Ceux qui nient que les externalistes sémantiques puissent retenir une quelconque notion de connaissance de soi sont des "incompatibilistes", tandis que ceux qui défendent la compatibilité de la connaissance de soi avec l'externalisme sémantique sont des "compatibilistes". Cet article examine les affirmations du compatibilisme, en se focalisant sur le "slow switching argument" de Burge et l' "objection d'alternatives pertinentes" de Boghossian. J'affirme que le compatibilisme est erroné et que l'externalisme sémantique est incompatible avec la connaissance de soi.

Mots-clés

Externalisme sémantique, anti-individualisme, compatibilisme, incompatibilisme, *slow switching*, connaissance de soi, épistémologie, philosophie de l'esprit