



4<sup>th</sup> International Conference of Cognitive Science (ICCS 2011)

## Slow switching and authority of self-knowledge

Hamed Bikaraan Behesht\*

*Philosophy of Science Department, Sharif University of Technology, Tehran, Iran*

---

### Abstract

Based on content externalism, the question of whether self-knowledge is authoritative or not has launched a real controversy in the philosophy of mind. Boghossian proposed slow switching argument in defense of incompatibility of the two views. This argument has been criticized by some philosophers through different approaches. Vahid is one of them. He claimed that Boghossian's argument appeals to some controversial assumptions without which it cannot achieve its conclusion. In this article, I criticize Vahid's response to slow switching argument and show that self-knowledge in Vahid's response differs in having authority and therefore in function from that of Boghossian.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of the 4th International Conference of Cognitive Science Open access under [CC BY-NC-ND license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

*Keywords:* Slow switching; self-knowledge; content externalism; a priority; Boghossian; Vahid

---

### 1. Introduction

According to content externalism, our thought-contents are not fully determined by internal or individual factors, hence, are dependent, at least partially, on environmental factors. This view seems to be in an apparent incompatibility with authoritative self-knowledge. Some philosophers have basically denied any incompatibility between content externalism and authoritative self-knowledge, and some others tried to argue in defense of the incompatibility. Slow switching argument, presented by Boghossian, is the main attempt to support the incompatibility thesis. This argument has been criticized by some philosophers through different approaches. Vahid is one of them. In Externalism, Slow Switching and Privileged Self-Knowledge (Vahid, 2003), he claimed that Boghossian's argument appeals to some controversial assumptions without which it cannot achieve its conclusion. So, he argued, if it could be shown that our self-knowledge was only negatively, and not positively, dependant on experience, then, based on a special view about a priority called modest a priorism, self-knowledge was still a priori. In this article, I criticize Vahid's response to slow switching argument and show that self-knowledge in Vahid's response differs in having authority and therefore in function – e.g., being used in an anti-skeptical argument – from that of Boghossian. It seems to me that self-knowledge should meet two conditions to be authoritative. First, it should be knowledge and, second, it should be gained independent of experience. But based on a relational – externalist – theory of mental content, there could be no idea about self-knowledge that bears both of these conditions for authority. Thus, what slow switching argument shows is that any view emphasizing on second condition lacks the first. Vahid, by focusing on the second condition while neglecting the first one, attempted to

---

\* Corresponding author. Tel.: +98-0261-4443396; fax: +0-000-000-0000  
*E-mail address:* [h.bikaraanbehesht@yahoo.com](mailto:h.bikaraanbehesht@yahoo.com)

disarm slow switching argument against compatibilism. If we pay further attention to the first condition, it can be shown that for self-knowledge to be knowledge, self-knowledge must be positively dependent on experience. Finally, I use the response of transcendental argument to skepticism to compare the role of the two types of self-knowledge, i.e., what Boghossian meant and what Vahid described. The result of the analyses offered in this paper is that slow switching argument rejected the authority of self-knowledge as accurately as possible and Vahid's criticism to this argument is not relevant.

## 2. Content externalism and self-knowledge

On the one hand, content externalism is a view about thought-contents based on which the content of one's thoughts, beliefs, desires and other propositional attitudes is determined not only by one's internal states, but also at least to some extent, by his external causal environment. The view was directly emerged by Burge (1979). He called his view *anti-individualism* and constructed it with aid of a well-known thought experiment that may be called Arthritis thought experiment. But before Burge, Putnam had introduced a similar view about the meaning of natural-kind terms (Putnam, 1975). He also introduced a well-known thought experiment to support his view that is known as Twin Earth thought experiment. Some philosophers like Burge and Brueckner have used Putnam's thought experiment to reestablish and argue for content externalism (Burge, 1982; Brueckner, 1990). These thought experiments proceed with the shared scenario that two persons who are molecule for molecule identical to each other, and who have same mental states *in narrow sense*, while inhabit in different physical or social environments, have different thought-contents and therefore different thoughts. First I turn to Arthritis thought experiment. Suppose Bert has several true beliefs about arthritis. For example, she believes that she has had arthritis for some years, or that arthritis is less dangerous than liver cancer, or that arthritis has several types. Besides these true beliefs, she has a false belief about arthritis, that her arthritis affected her thigh, while arthritis is inflammation of joints and does not occur in thighs. Hence if Bert goes to a physician, the physician will tell her that she makes a mistake about her thigh's disease. Now suppose a counterfactual situation in which Bert has same experiences as actual case before visiting the doctor, and her physical history, internal events and sensory perceptions from birth to the time she visits the doctor are same to the actual case. The only difference is in her society that the word 'arthritis' has different meaning from actual case. In this counterfactual society, 'arthritis' means arthritis plus the disease of Bert's thigh. This imaginary concept called tharthritus. In this counterfactual situation, if Bert says "I have arthritis in my thigh", she says a true statement and her belief relevant to this statement is a true one. Therefore, without any changes in actual Bert and counterfactual Bert, and with only a change in the environments of them, the thought-content of Bert has changed.

As noted above, Putnam's Twin Earth thought experiment is also used to support content externalism. Suppose a planet somewhere within our galaxy that is similar to earth. Putnam called it twin earth. Everything in earth has a duplicate in twin earth. Twin earth is completely same as earth except the difference that the liquid in oceans, seas, lakes and rivers of twin earth and falling as rain from its sky is not water, but is a liquid with same superficial characteristics as water that has a different chemical formula instead of H<sub>2</sub>O, abbreviated as XYZ. XYZ is not distinguishable from water without chemical experiments requiring laboratory equipment. This imaginary liquid is called twater. The condition of the thought experiment is such that most of inhabitants of earth and twin earth are not aware of chemical formula of what they call 'water'. For example, Putnam's story about twin earth occurs in 1750, the time that water's formula is unknown to scientists. Now suppose Oscar, an inhabitant of earth, has a duplicate, Toscar, on twin earth. When Oscar says "I believe that water is wet", Toscar also says that statement on twin earth. But the content of Oscar's belief is water is wet, and the content of Toscar's belief is twater is wet. Because the liquid that Toscar is in contact with is twater not water. So when these two persons are physically the same, they have different thought-contents, because of a difference in their environments. Boghossian called this view *relationalism* and pointed out that in this view "the content properties of mental states and events are determined by, or supervenient upon, their relational properties" (Boghossian, 1989, p. 157) and that "many of a person's thought contents are necessarily dependent on relations that that person bears to the physical or, in some cases, social environment" (p. 157). Despite various criticisms that have been launched against this view and its supporting thought experiments, still its central core is very plausible.

On the other hand, there is an intuitive conviction that we have a *privileged access* to our thoughts and we know directly and without any mediator the content of our thoughts. That is, we benefit an authoritative self-knowledge.

This view is highly intuitive and as Davidson said, "'usually' it doesn't make sense to ask someone why he believes he has a certain belief or desire or intention" (Davidson, 1984, p. 103). But it is important to note that the person's direct access to his thoughts is only acceptable in the realm of his mind. It is necessary to use sense for knowing about environment. Therefore, one has not privileged access to his external environment. So content externalism and authority of self-knowledge seem to be in an initial contrary. According to authority of self-knowledge, one's direct knowledge to his thought-contents is merely realized in the realm of his mind and not in external environment. However, based on content externalism, we can say that the content of one's thoughts, at least to some extent, is lodged in the external environment. Consequently, one could not know directly his thoughts' content. Boghossian presents this point by a question: "[H]ow could anyone be in a position to know his thoughts merely by observing them, if facts about their content are determined by their relational properties?" (Boghossian, 1989, p. 157).

The issue proceeded with rich amount of analyses. Some may accept this contrary and therefore believe that content externalism is incompatible with authoritative self-knowledge; the *incompatibilists*. One of the most important attempts to support incompatibilism is slow switching argument posed by Boghossian (1989). But some others have attempted to show the compatibility between content externalism and authoritative self-knowledge; the *compatibilists*. They attempt to criticize and undermine the slow switching argument. Vahid may be one of them. Although, he himself criticized some responses to that argument, but he tried to launch a strike against the main argument of the incompatibilists. In the next section, slow switching argument will be described.

### 3. Slow switching argument

As noted in previous section, content externalism and authoritative self-knowledge initially seem to be incompatible. This incompatibility will be better determined if we refer to Twin Earth thought experiment. In that thought experiment, the content of Oscar's and Toscar's beliefs is determined in relation to the liquid that is called 'water' in their environments. So it seems if they want to know what they are thinking about, water or twater, they must observe their environments, and without it and by mere introspection they could not know which one is the content of their thoughts. This initial incompatibility has been blocked with the aid of a special view about knowledge: the theory of *relevant alternatives*. According to Boghossian, this view is "[t]he ordinary concept of knowledge" (Boghossian, 1989, p. 158). Based on this view about knowledge, for knowing something, it is not necessary to know all other possible alternatives are not established. This view also could be applied to self-knowledge. When one says that he believes "water is wet", the thought twater is wet could be a threatening alternative thought to the water-thought, if that thought be a relevant alternative to water-thought. But a twater-thought in normal condition is not a relevant alternative to a water-thought. Therefore, the guy knows what thought he has. This point is firstly emphasized by Burge (1988, pp. 654-655), and Boghossian referred to it (Boghossian, 1989, p. 158).

Slow switching argument could be seen as an attempt to present a scenario in which twater-thought be a relevant alternative to water-thought. Boghossian's scenario combines the thought experiments of Putnam and Burge. Suppose S, a person on earth, undergoes a series of switches between earth and twin earth and these switches are occurred such that S is not aware of them. S remains on every planet long enough to be said his concepts is *slowly changed* from a place to another. For example, when he says "arthritis is painful" on earth, his statement is about arthritis, and when, after enough dwelling on twin earth, he says the same statement, his statement will be about tharthrititis. After these switches, when he comes back to his home on earth, it could be said that when he has a thought in connection to the word 'arthritis', tharthrititis-thought will be a relevant alternative to his arthritis-thought, while the content of his thought is indeed arthritis. Boghossian truly emphasized that for this relevance, S should not be aware of it, for "[e]pistemic relevance is not a subjective concept" (p. 160). If S is asked whether he has recently thought about an arthritis-like concept, he will respond "no". This negative response is not because he has forgotten his tharthrititis-thoughts, but it is because he never knew them (pp. 159-160). Therefore S, like other people, do not benefit from authoritative self-knowledge. In the next section, I will describe Vahid's criticism of this argument.

### 4. Vahid's criticism of slow switching argument

Vahid criticized Boghossian's argument (Vahid, 2003). He saw the result of slow switching argument as showing that our knowledge to our thought-contents is susceptible to empirical contingencies. But the concept susceptibility

to empirical contingencies is vague. Because, this susceptibility may be positive or negative. For determining that the susceptibility to empirical contingencies is of which type, it should be determined that how our self-knowledge is dependent on experience for its justification (p. 379). In fact, Vahid has focused on the method with which one knows his own thought-contents. He said: "Moreover, the issue here is not whether one can know what thought one is having, if the thought is individuated by environmental factors. The issue is whether one can know in a privileged way if the thought is individuated by environmental factors" (p. 379). As is obvious, for Vahid, it is important not only that we can know our thoughts, but also how we reach to that knowledge.

A belief is positively dependent on experience of the environment if and only if the experience of the environment plays an important role in producing or supporting the justification of that second-order belief. And a belief is negatively dependent on experience of the environment if and only if that justification will be undermined with empirical evidence. But what sort of dependency is in self-knowledge? Vahid argued that the justification of our self-knowledge is not such that it could be said that our self-knowledge is positively dependent on experience. Because when one tells about his belief, if he is in a normal condition and his cognitive power is not impaired, then he is justified in his belief. Also his justification is a result of reflective thought and therefore is non-empirical. So, we can say that self-knowledge is not positively dependent on experience of the environment. But is self-knowledge negatively dependent on experience? Vahid answered this question with "yes". He believed that it could be seemed a special condition in which one wrongly comes to believe that he has not gotten knowledge toward a thought of him, based on experience of environment. Vahid's example is a person that some reputable philosophers *misleadingly* convinced him that he is not thinking about water; however he is himself suspicious that he is. These philosophers based their claim on reading EEG of that person.<sup>†</sup> In such a case, empirical evidence is used to overthrow a second-order belief. (pp. 380-381)

Now, the question here is that if self-knowledge is negatively and not positively dependent on experience, then does self-knowledge fall from a priority? Following a special view about a priority, called *modest a priorism*, Vahid answered negative response to this question. According to modest a priorism, a proposition is a priori, if it is not positively dependent on experience. Negative dependence on experience is not a threat for a priority of a proposition in this sense. So slow switching argument could proceed only with the presupposition that a proposition, to be a priori, needs to be both negatively and positively independent from experience. This is probably the controversial assumption Vahid noted in the beginning of his paper. (pp. 381-382)

## 5. Authority of self-knowledge

For assessing Vahid's criticism of slow switching argument, it is necessary to go through what is called "self-knowledge" in more details. It could be said that from self-knowledge, Boghossian meant authoritative self-knowledge. But what is authoritative self-knowledge? I refer to a statement from Goldberg that seems to be an acceptable definition for authoritative self-knowledge. He stated that "[t]he doctrine of authoritative self-knowledge of content is the view that, for any thinker S and occurrent thought that p, S can be said to have a priori knowledge of the content of her thought that p" (Goldberg, 2000, p. 52), in which "a priori knowledge' is roughly equivalent to 'knowledge whose justification is not dependent on empirical investigation'" (p. 75, n. 2). According to this definition authority of self-knowledge has two conditions. That is, for self-knowledge to be authoritative, it must have the following two conditions:

- I) Self-knowledge should be knowledge.
- II) Self-knowledge should be a priori, that is, its justification should be independent of empirical investigations.

But did Boghossian really mean this concept of self-knowledge? There is some evidence that he did. He said in the introduction of his paper that "[a] couple of preliminary remarks before we proceed. First, I propose to be reasonably serious in the use of the term 'knowledge': by 'self-knowledge' I shall mean not just a true belief about

---

<sup>†</sup> There is a key point here. From Vahid's point of view, the process of convincing aforementioned person is a misleading process, not a non-misleading one. Also he believes that EEG could not be used to overrun a person's self-knowledge (p. 381, n. 28).

one's own thoughts, but a justified one" (Boghossian, 1989, p. 150)<sup>‡</sup>. It is apparent that being knowledge is a condition for being self-knowledge. Albeit Boghossian, declared that he did not want to be so serious to enter into Gettier's challenge of definition of knowledge. But what is certain is that he believed knowledge as indispensable element of self-knowledge – authoritative self-knowledge in this context.

Boghossian also considered what the condition (II) considers as necessary for authoritative self-knowledge. From Boghossian's point of view, self-knowledge has a direct character. He declared that "[t]he intuitive epistemic facts indicate that knowledge of one's mental states is direct. ... We may conclude, on the one hand, that self-knowledge is not inferential because it is based on some form of inner observation; or, on the other, that is not inferential because it is based on nothing – at any rate, on nothing empirical" (p. 156). Directness of self-knowledge that was noted by Boghossian, tells us about the method with which we come to know our thoughts, which is its justification. To be based on inner observation or nothing – nothing empirical – results in that self-knowledge is in no way dependent on experience, and so we can say it is a *priory*. The target of Boghossian's argument to attack is that self-knowledge is based on inner observation. So the self-knowledge meant by him, has shared the second condition of authoritative self-knowledge as noted.

Therefore Boghossian (1989) aimed a concept of self-knowledge presented here. He noted implicitly these two conditions in some other places of his paper. It is important to note that Boghossian did not attribute infallibility and unrevisability to self-knowledge. He pointed to Descartes' view about self-knowledge that raised self-knowledge with infallibility and exhaustivity, and declared about these claims that "for a certain restricted class of mental events – namely, sensations – they may even be true" (p. 151). But he rejects Cartesian claims in the case of "contentful or representational states", that is the case in self-knowledge (p. 151).

## 6. A closer view to Vahid's criticism

We could construe Boghossian's strategy in slow switching argument as he tried to show that in the light of content externalism, conditions (I) and (II) about authoritative self-knowledge could not be complied together. He showed that if we emphasize on condition (II), condition (I) could not be saved. He asked a question about this: "Now, doesn't it follow from such anti-individualistic views that we cannot know our thoughts in a direct, purely observational manner?" (Boghossian, 1989, p. 158). But, as could be seen from above, Vahid's emphasis was on condition (II). Because he attempted to show that Boghossian's thought experiment finally could not threaten the priority of self-knowledge. Now if Vahid in his analyses in defense of the establishment of condition (II) in self-knowledge has saved the condition (I) for self-knowledge, then it could be said that he has successfully rejected the Boghossian's argument. But it does not seem to me to be so. For showing this point, I will carry out two considerations.

Firstly, Vahid accepted that self-knowledge is negatively dependent on experience. But the dependency Vahid intended cannot destroy self-knowledge. Vahid's example for this dependency, as said earlier, was a person that *misleadingly* convinced by some reputable philosophers that he doesn't have a water-thought. But is it impossible to show a condition in which an evidence *non-misleadingly* convinces, or better to say, shows a person that he was in error in a self-knowledge of him? If it could be, then, in fact, that person has a false second-order belief, not a true one. So the condition (I) for being authoritative self-knowledge will be marred. It seems to me that such condition could be satisfied. I offer a thought experiment in which I use the twin earth scenario to describe such a condition.

Suppose a period that the chemical formula of water had not been discovered. Yet, a chemist in that time is studying on water and discovers some important features of water with a series of experiments. His experiments on water have a specified set of results, I call it R1. Within his experiments, for example, he electrolyzes some water and gets some oxygen. After these experiments, he stops his scientific activities and goes to a seaside for vacation. In first night of his vacation, without being aware, someone transfers him from earth to twin earth. He stays there for a month and in that time period he deals with twater for his usual needs. So it seems we can say his thought-content about what is called "water" in both earth and twin earth, slowly switches from water to twater. After his vacation finishes, he comes back to his laboratory, not to his real laboratory in earth, but to its counterpart in twin earth. There, he repeats once more his experiments and observes that his experiments do not lead to R1 results, but to some

---

<sup>‡</sup> Vahid also referred to this statement of Boghossian (Vahid, 2003, p. 378).



other result set, R2. For example, when he electrolyzes twater, he does not get oxygen. He repeats his experiments again and again, and gains same results. So he at last comes to believe that he is in error about the liquid he washed his hands with, before he started his experiments. In such condition, a belief was truly undermined with some empirical evidence.

Secondly, even a situation could be described in which a person needs to observe his environment to know his thought-content, if we want to preserve condition (I) for authoritative self-knowledge. In other words, with emphasizing on condition (I), condition (II) would be marred. Now suppose after those events occurred for our chemist, every night, he is transferred between earth and twin earth; one night from earth to twin earth, and the next vice versa. So he would be doubtful about the liquid within his hands, because a day his experiments lead to the result set R1, and next day they lead to the result set R2. So every time he wants to know which liquid is in front of him and he is thinking about, water with R1 results or twater with R2 results, he should do his experiments again. In fact, he has two concepts linked to liquids with same superficial characteristics, and he experiences the nature such that each day he is confronted with one of them and not both.<sup>§</sup> So if we emphasize on condition (I), then we miss the condition (II).

The concept of self-knowledge that we can extract from Vahid's response to slow switching argument is different from the concept of the authoritative self-knowledge. Vahid himself attended to the conditions (I) and (II) in his paper, especially in his criticism of *standard strategy* response to slow switching argument. But when he was proposing his own response, the picture of self-knowledge depicted in his analyses seems to ignore condition (I).

## 7. Self-knowledge in transcendental argument

Some of the compatibilists believe that not only content externalism is compatible with authoritative self-knowledge, but these two views together constitute an argument against skepticism. In this section I will review the transcendental argument against skepticism, in which authoritative self-knowledge is used. I will try to show that the concept of self-knowledge in Vahid's response cannot be used instead of authoritative self-knowledge in that argument. Then, we can say that these two concepts of self-knowledge are different in function.

First, I will look at the transcendental argument. This argument was proposed by Brueckner (2003) to solve the problems of Putnam's anti-skeptical semantic argument. This argument targets proving that I am not a BIV. Based on content externalism, if I am a BIV, then my thought-contents is determined in connection with objects that I am in relation with. For example, if I am a BIV and say "I believe that trees are green", then my thought-content is not about trees, but is about tree\**s*. Simply we can say that tree\* is a computer program feature that produces the electronic impulses of a BIV's brain and leads to the tree-image that I perceive. Transcendental argument benefits from two premises; one from content externalism, and other from authority of self-knowledge. Here is a version of that argument:

1. If I am a BIV, then I am not thinking that water is wet.
2. I am thinking that water is wet.
- ∴ I am not a BIV.

The first premise of this argument comes from content externalism, as described above, and its second premise is a result of authoritative self-knowledge. If we have authoritative self-knowledge, then we know our thoughts and our thought-contents without appealing to experience. In case of a BIV, it is not possible to investigate the nature – external world – to find out our thought-content. So, the second premise is very crucial for this argument. The second premise implicitly implies that I think that water is wet and I do not think that water\* is wet. This argument is from first-person's point of view, so we can say that the second premise results in the premise that I know that I do not think that water\* is wet. Now this is the question: is it possible to replace the concept of self-knowledge in Vahid's response with authoritative self-knowledge in transcendental argument? The justification that Vahid posed

---

<sup>§</sup> The chemist was not aware of his transfers between earth and twin earth. So he may think that the nature changes every day.

for self-knowledge is based on three things. Vahid said that a person on earth knows that he has a water-thought because, first, his first-order belief about water – his belief that water is wet – is justified since the liquid in front of him has all superficial characteristics of water. Second, his cognitive power is not impaired and the circumstances under which he forms his belief is normal. And third, his second-order belief is a result of a reflexive thought, so it is non-empirical. But I do not think that with such justification one can claim that he knows that he is thinking that water is wet and not thinking that water\* is wet. Because if I am a BIV, similar circumstances exist: the liquid in front of me has same superficial characteristics with the liquid that we call 'water'. If some neuroscientists, or better to say, neuroscientist\*s consider my cognitive system, they will approve that my nervous system is not impaired. And my second-order belief is result of a reflexive thought, so it is not empirical. But still I am a BIV thinking about water\*. Therefore, Vahid has saved the a priority of self-knowledge at the expense of weakening self-knowledge so that it could not be used in an important application like transcendental argument.

## 8. Conclusion

According to what was presented above, the concept of self-knowledge through Vahid's response to slow switching argument differed from that of Boghossian. So what Vahid criticized was not exactly what Boghossian meant. Besides, by weakening the concept of self-knowledge, Vahid sacrificed the role of self-knowledge in transcendental argument, because authoritative self-knowledge was too strong to be replaced with a notion like the one used by Vahid in his response to that argument.

## References

- Boghossian, P. A. (1989). Content and self-knowledge. In P. Ludlow, & N. Martin (Eds.), *Externalism and self-knowledge* (pp. 149-173). Stanford: CSLI Publications.
- Brueckner, A. (1990). Scepticism about knowledge of content. *Mind*, 99, 447-451.
- Brueckner, A. (2003). Trees, computer program features and skeptical hypotheses. In S. Luper (Ed.), *The skeptics: Contemporary essays* (pp. 217-226). Burlington: Ashgate.
- Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy*, 4, 73-121.
- Burge, T. (1982). Two thought experiments reviewed. *Notre Dame Journal of Formal Logic*, 23, 284-293.
- Burge, T. (1988). Individualism and self-knowledge. *The Journal of Philosophy*, 85, 649-663.
- Davidson, D. (1984). First person authority. *Diabetica*, 38, 101-111.
- Goldberg, S. (2000). Externalism and authoritative knowledge of content: A new incompatibilist strategy. *Philosophical Studies*, 100, 51-79.
- Putnam, H. (1975). The meaning of 'meaning'. In H. Putnam (Ed.), *Mind, language and reality: Philosophical papers* (pp. 215-271). Cambridge: Cambridge University Press.
- Vahid, H. (2003). Externalism, slow switching and privileged self-knowledge. *Philosophy and Phenomenological Research*, 66, 370-388.