

CRISTINA BORGONI

WHEN EXTERNALISM AND PRIVILEGED
SELF-KNOWLEDGE ARE COMPATIBLE
AND WHEN THEY ARE NOT*

Abstract: This paper is dedicated to the incompatibilist debate between externalism and privileged self-knowledge, such as it appears in the literature under two privileged contexts of discussion: the slow-switching cases and the *reductio ad absurdum* arguments. My aim is to defend a compatibilist position although recognising some exceptions to it. I will defend, on the one hand, that the incompatibilism reached by slow-switching cases is sustained only in case we maintain a specific but problematical view about self-knowledge. On the other hand, the incompatibilism reached by *reductio ad absurdum* arguments is only sustained if we maintain a narrow conception of externalism.

Keywords: incompatibilism, slow-switching cases, *reductio* of compatibilism.

CUANDO EL EXTERNALISMO Y
EL AUTO-CONOCIMIENTO PRIVILEGIADO
SON COMPATIBLES O NO

Resumen: Este trabajo se dedica al debate incompatibilista entre externalismo y autoconocimiento privilegiado, tal como aparece en la literatura bajo dos contextos específicos de discusión: los casos de transferencia lenta [slow-switching cases] y los argumentos de tipo *reductio ad absurdum*. Mi objetivo es defender una posición compatibilista que a la vez reconozca algunas excepciones a ella. Defenderé, por un lado, que los casos de transferencia lenta logran dar bases a un incompatibilismo solamente si mantenemos una visión específica pero

Recibido 21-11-08 ✧ Aceptado 27-11-08

* I would like to thank Manuel de Pinedo and Sarah Sawyer for comments and suggestions on all aspects of this paper. I am also grateful to Jesús Palomo Muñoz and César Marini. This work has been partially supported by the research project HUM2007-63797/FISO (MEC).

problemática acerca del autoconocimiento. Por otro lado, los argumentos de tipo *reductio ad absurdum* logran dar bases a un incompatibilismo solamente si mantenemos una concepción estrecha del externalismo.

Palabras clave: incompatibilismo, casos de transferencia lenta, *reductio* del compatibilismo.

1. Introduction

As Ludlow¹ suggests, externalism is in a sense “the denial of the traditional Cartesian view that holds that the contents of our thoughts are what they are independently of the surrounding world”. What philosophers normally take to be Cartesianism² is not only committed to internalism about mental content, but mainly to a view on self-knowledge according to which such a realm plays a fundamental role both in epistemology and in metaphysics. Under this approach, self-knowledge is conceived as a kind of knowledge entirely acquired by privileged means, that is, in a way that dispenses with any empirical investigation and with any inferential process.

Following this reasoning, once externalism is incompatible with Cartesianism, one should expect externalism to be incompatible also with the possibility of privileged self-

¹ Cf. Ludlow, P., “Externalism, Self-Knowledge, and the Prevalence of Slow Switching” (1995a) in Ludlow, P. & Martin, N. (eds.), *Externalism and Self-knowledge*, Stanford, CSLI Publications, 1998, p.1.

² After Burge defended that Cartesianism was committed to internalism, or better, that “Individualism as a theory of mind derives from Descartes” (Burge, T., “Cartesian Error and the Objectivity of Perception”, 1986 in *Foundations of Mind*, Oxford, Clarendon press, 2007, p. 192), he portrayed himself as a defender of the anti-individualism spirit in Descartes’s works (Burge, T., “Descartes on Anti-individualism” (2003/2006) in *Foundations of Mind*, Oxford, Clarendon press, 2007, pp. 420-439.). This seems to be an open question, but as far as a kind of established caricature called ‘Cartesianism’ exists -a position committed both to individualism about mental contents and to self-knowledge entirely acquired by direct and non-empirical means- I will refer to such a theoretical position when talking about Cartesianism.

knowledge³: one thesis or the other obtains, but not both. However, as we all know, Cartesianism is neither the only nor the best available account of self-knowledge, even if we don't forfeit its special trait such as its acquisition by us in a direct and non-empirical manner.

This paper is dedicated to the incompatibilist debate between externalism and privileged self-knowledge, such as it appears in the literature under two favoured contexts of discussion: the slow-switching cases and the *reductio ad absurdum* arguments⁴. My aim is to defend a compatibilist position although recognising some exceptions to it. I will defend, on the one hand, that the incompatibilism reached by slow-switching cases is sustained only in case we maintain a specific but problematical view about self-knowledge. On the other hand, the incompatibilism reached by *reductio ad absurdum* arguments is only sustained if we maintain a narrow conception of externalism.

In the first part, I shall discuss some incompatibilist arguments based on the thought experiment of slow-switching and their respective compatibilist answers. I will also discuss the role such thought experiments play in the general context of the discussion, indicating that a compatibilist answer could be designed without the consideration of such cases. At this point, I will compare Tyler Burge's and Donald Davidson's compatibilism.

In the second part, I will treat the reduction arguments. I will defend that the externalist premise over which those arguments are constructed is misleading in relation to what an externalist is committed to.

2. *The slow-switching cases*

³ The term 'privileged self-knowledge' will be used here in reference to the direct and non-empirical way by which we acquire at least part of our self-knowledge. It is important to notice that both aspects – directness and non-empiricism – will be required to characterize such a specific knowledge.

⁴ Respectively referred by Davies, M., "Externalism and Armchair Knowledge" in Boghossian, P. & Peacocke, C. (eds.), *New Essays on the A Priori*, Oxford, Oxford University Press, 2000, p. 391, as 'the achievement problem' and 'the consequence problem'.

The thought experiment that establishes the first discussion context is exposed by Burge in 1988, in a paper where he defends a compatibilist position. It is the so called ‘thought experiment of slow-switching cases’, where a subject - let’s say Oscar - is stealthily shifted back and forth between actual Earth and Twin Earth, several times, remaining unaware of those shifts. Oscar acquires the appropriate concepts to each situation, such as water and twater (twin water). If Oscar is told about such switches and asked to identify when they took place, he will not be able to answer.

Boghossian’s⁵ comments (1989) on slow-switching cases have given rise to two different incompatibilist arguments: one that emphasizes the discrimination of mental contents from their relevant alternatives and another one which emphasizes the question about memory.

2.2. *Discrimination of mental contents and relevant alternatives*

The first of those incompatibilist arguments can be restructured 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 of water thoughts.
- (C1) So, S doesn’t know that P by introspection⁶.

This argument stresses the consequences of being unable to distinguish between actual and twin situations. The underlying intuition is that in order to have knowledge of any content, one should be able to distinguish it from the relevant possibilities. The thoughts Oscar has on Twin Earth establish relevant alternatives to

⁵ Boghossian, P., “Content and Self-Knowledge”, in Ludlow & Martin, (eds.), *Externalism and Self-knowledge...*, cit., pp. 149-173.

⁶ Warfield, T., “Privileged Self-Knowledge and Externalism are Compatible”, in Ludlow & Martin, (eds.), *Externalism and Self-knowledge...*, cit., p. 218.

the thoughts he has on Earth. However, Oscar is unable to distinguish between them only by introspection.

On the basis of such an argument there is an important distinction between merely logical alternatives to one's thoughts and relevant ones. In the standard externalists' scenarios, such as the one proposed by Putnam⁷, T. Oscar's (Twin Oscar's) thoughts about twater represent only logical alternatives to Oscar's thoughts about water. In those cases, to require Oscar to discriminate his water thoughts from twater thoughts would establish an implausible condition to knowledge, such as discriminating one thought from every single alternative possibility to it. However, in slow-switching cases, insofar as the subject of the switches seems to have both concepts⁸, to discriminate between them seems to be more acceptable.

As Boghossian points out,

the ordinary concept of knowledge appears to call for no more than the exclusion of 'relevant' alternative hypotheses [...] and mere logical possibility does not confer such relevance⁹.

In order to know that I have €2,25 in my pocket, I do not need to have checked that there is no forgery money in the vicinity, nor do I need to be able to tell the difference between a genuine euro and every imaginable forgery to that¹⁰. I just have to count the coins. But, if I had 20p together with my Euros, I should be able not to count them. In this case, differentiating Euros from Pounds seems to matter to my final knowledge.

Following Boghossian's argument, Oscar would not have knowledge by introspection of his own thoughts about water because he is unable to distinguish them, also by introspection, from his other thoughts about twater. It seems that in order to distinguish them –and so, to know them– Oscar would have to

⁷ Putnam, H., "The Meaning of 'Meaning'" (1975) in *Mind, Language and Reality*, Cambridge, Cambridge University Press, 1975, pp. 215-271.

⁸ There are important nuances in the interpretation of the case that the subject possesses two concepts. This question will arise later on in this paper.

⁹ Cf. Boghossian, P., "Content and Self-Knowledge", cit., p. 158.

¹⁰ Cf. *Ibidem*.

engage in an empirical search, which will be favourable to incompatibilism.

However, this first version of Boghossian's comments has no effect at all on Burge's compatibilist strategy developed in 1988. His position consists in showing that there is a class of self-knowledge, named as the 'basic' one, which would resist such proofs. We would not need to differentiate the items of this class from their relevant alternatives because this group has the characteristic of being self-verifying. In this case, (P1) would be false¹¹.

Despite the inability to discriminate between twin periods from home ones, Burge will argue, the subject of the experiment is still able to have privileged self-knowledge, at least in reference to the so called '*cogito*-like judgements': a range of second-order thoughts which are "self-verifying" because of their self-referential form, such as "I think that I am thinking that water is wet".

The appeal to the *cogito*-like judgments guarantees that, at least in a specific range of self-knowledge, we can find not only the externalist aspect of mental contents, but also a special way to acquire them. Burge's position is, in fact, somewhat stronger than that, because he takes *cogito*-like judgments to be the paradigmatic instances of self-knowledge. That is why he identifies them as the 'basic self-knowledge'. Although this class of knowledge plays such an important role in Burge's position, it is crucial to notice that, according to him, not all self-ascription of beliefs are self-referential or self-verified. He maintains that a variety of self-knowledge cases extend out of what he has called the basic one¹². However, this special class has some epistemic peculiarities. According to Burge¹³:

¹¹ Furthermore, (P1) is false in Burge's account because, according to him, discriminating between relevant alternatives plays a more decisive role in empirical judgments than it does in self-knowledge.

¹² Cf. Sawyer, S., "In Defense of Burge's Thesis", en *Philosophical Studies*, vol. 107, 2002, p. 112. Sawyer points out that Burge's thesis doesn't constitute a general theory of authoritative self-knowledge. Nevertheless this is not necessary for providing a good compatibilist answer.

¹³ Burge, T., "Individualism and Self-Knowledge", en *The Journal of Philosophy*, vol. 85, 1988, 11, p. 660.

The source of our strong epistemic right, our justification, in our basic self-knowledge is not that we know a lot about each thought we know we have. It is not that we can explicate its nature and its enabling conditions. It is that we are in the position of thinking those thoughts in the second-order, self-verifying way. Justification lies not in the having of supplemental background knowledge, but in the character and function of the self-evaluating judgments.

Burge insists that this specific group of second-order thoughts are self-referential, and hence self-verified, partly because the first-order thought, externalistically individuated, is somehow embedded in the second-order thoughts. This idea has constituted the most accepted answer to this first formulation of incompatibilist worries. According to Davies¹⁴, in order to sustain compatibilism, several positions have based his answer on the fact that:

[W]hen I think that I am thinking that water is wet, I deploy in thought the very same concepts of water and of being wet that are involved in my thinking that water is wet. So an externalist dependence thesis that is true for my first-order thinking that water is wet will be no less true for my second-order thinking that I am thinking that water is wet. Because the content of my second-order thought embeds the content of my first-order thought, my second-order thinking shares the dependence on the environment that is characteristic of my first-order thinking.

This point will be discussed again in the following sections, but one thing is important to retain. The appeal to this kind of condition to the second-order thoughts does not exactly mean following Burge in his answer to the incompatibilist problem. His answer is quite stronger insofar as it is sustained by the conception of a specific self-knowledge class, the basic one¹⁵.

¹⁴ Davies, M., "Externalism and Armchair Knowledge", in Boghossian & Peacocke (eds.), *New Essays on the A Priori*, cit., p.391.

¹⁵ We could indeed see Burge's proposal as composed by two elements: the reference to basic self-knowledge (which offers an evidence that there is self-knowledge in an externalist scenario) and the reference to the "embedding condition" (which explains that the concept employed in the second-order

As a result of this, one could insist that the restricted group to which Burge refers doesn't satisfy incompatibilist worries. It would be necessary to talk about self-knowledge in general. More than that, basic self-knowledge could not even resemble what we would like to take as the representative group of self-knowledge. So, just as a theoretical device, let's exclude Burge's strategy for a while in order to understand a little more about the core of this sort of incompatibilist argument.

Ludlow¹⁶ not only considers Boghossian's argument (as restructured by Warfield) to be a cogent one, but also proposes a stronger reading of it. He adds the following premise to the argument above:

“(P4) Switching cases, in general, are prevalent”¹⁷.

Ludlow claims that because “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”¹⁸, we are subject to situations very similar to those proposed by slow switching cases. He maintains that departing from what he identifies as ‘social externalism’ – “namely that content is socially determined and that the relevant social groups may be highly localized”¹⁹ – premise (P4) is entirely plausible. In doing so, Ludlow considers that Boghossian's argument can be better defended.

Ludlow exemplifies his position with the English word ‘chicory’, which designates two different but seemingly similar vegetables in England and United States. And he imagines a British traveller who constantly goes from one country to the other. In addition, the traveller remains long enough in the United States and consequently acquires the mental content related to that environment, in such a way that the traveller would have his thoughts shifted each time he enters each country, remaining however unaware of this.

thought is the same as the one employed in a first-order thought that is individuated by externalist conditions).

¹⁶ Ludlow, “Externalism, Self-Knowledge ..., cit., p.227.

¹⁷ *Ibidem*.

¹⁸ *Ibid.*, p. 228.

¹⁹ *Ibid.*, p. 229.

Although Ludlow²⁰ presents an interesting line of argument when he brings thought experiments to our daily lives, it seems that his thesis about the prevalence of those cases is misleading. For it seems that, in order to really imagine a situation where a person remains unaware of the double aspect of a word while she uses it constantly, we would have to imagine her completely isolated of any social contact. Let's think of another word, for example, 'chips', which in Britain means strips of potato fried in deep fat (which Americans call 'French fries'), while in the United States, the land of poker, it means casino tokens. It is very unlikely that a person –Carol, for example– coming from the United States, might use such a word without originating an initially conflictive situation that could be easily solved at a certain moment. Whenever she hears something like "these chips are delicious with vinegar", she would inevitably learn the second use of the same word.

In this way, the fact that there are some daily situations where we may find similarities to slow-switching cases doesn't mean that we are subject to them most of the time. What is more, the requirement of remaining long enough in the other environment in order to acquire the mental contents related to such a place is not a mere question of passage of time. It refers exactly to the fact that while the subject was there, there were interactions between her and the objects of such an environment, as well as between her and the people from that place. The problem seems to be that Ludlow assumes that the subject could be maintained inert to such interactions. Externalism is, however, exactly the opposite of such an idea. Those interactions constitute the very subject and her new experiences.

Considering the way we learn new words, it seems that such learning includes indeed knowledge about the different contexts where they are used. That is why, several times, we prepare ourselves not to use some words or expressions when we are going to other countries, other cities or even to other social contexts. We usually know that they have other meanings elsewhere. Furthermore, we normally realise when we are entering a different

²⁰ *Ibidem.*

social group or engaging in a different language game. Externalism doesn't require our ignorance about different environments.

Let's go back to Boghossian's argument in its original version. It seems that an important criticism is the one developed by Warfield²¹.

Warfield criticizes such an argument stating that

Boghossian's shows at most that those individuals who are being slow switched fail to know the contents of some of their thoughts²².

That is, conceding some extent of soundness to Boghossian's argument, it doesn't go any further than showing that "externalism is *consistent with* a lack of self-knowledge; it does not show that externalism *implies* a lack of self-knowledge"²³.

According to Warfield²⁴:

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. Boghossian shows at most that some possible worlds are worlds in which externalism is true and individuals lack privileged self-knowledge and Ludlow [1995a] shows at most that one world, the actual world, is a world in which externalism is true and (some) individuals lack privileged self-knowledge.²⁵ [my italics].

²¹ Warfield, "Privileged Self-Knowledge and...", cit.; Warfield, "Externalism, Privileged Self-knowledge, and the Irrelevance of Slow Switching" (1997), in Ludlow, & Martin, (eds.), *Externalism and Self-knowledge*, cit., 1998, pp. 215-221.

²² Cf. Warfield, "Externalism, Privileged Self-knowledge...", cit., p. 232.

²³ *Ibidem*.

²⁴ Cf. *Ibid.*, p.233.

²⁵ Ludlow understands this condition as follows: "Warfield's insistence that I show privileged self-knowledge to be false in every possible world in which externalism is true completely inverts the argumentative burden here." (Ludlow, *ibid.*, p. 236). I would insist that the burden of the proof is in fact with the incompatibilist because, on the one hand, slow-switching cases are not but abnormal situations, and by other side, one cannot reach incompatibilism from those cases unless one defends a specific approach of self-knowledge.

Even in the stronger version defended by Ludlow –which I have considered as misleading– incompatibilism would not hold for similar reasons. Ludlow’s conclusion was that most of us are most of the time under slow-switching cases and because of that, most of us fail to know our own thoughts if we take externalism to be true. But the conclusion required in order to reach incompatibilism would be that, considering externalism as true, all of us fail to know by privileged means every single thought we have.

At this moment, the impression is that the argument doesn’t fulfil Boghossian’s or Ludlow’s expectations. Showing some cases where externalism is taken as true while individuals lack self-knowledge doesn’t prove anything else than a compatibilism between externalism and a lack of self-knowledge; a conclusion not only acceptable, but quite accurate. It seems that there is enough data favouring the idea that we lack such an easy knowledge about all our thoughts. Failures of self-knowledge such as self-deception and akrasia seem to be merely the extreme cases that corroborate such an idea. On the one hand, privileged self-knowledge seems to be plainly true, but on the other hand it seems that we need to recognize that some range of self-knowledge is acquired by other manners than the privileged one.

In that sense, there seems to exist one situation where Boghossian’s incompatibilism would work: if one insisted that there could not be a case of self-knowledge which was not potentially knowable to the subject in a direct and non-empirical manner. In slow-switching cases, self-knowledge about one’s water thought may fail at a certain moment.

The fact that Oscar is unable to discriminate between his water thoughts and his twater thoughts is not likely to affect *cogito*-like judgments. If Oscar states “I’m thinking that I think that fish live in the water”, he is probably right about what he is thinking. However, it can affect self-knowledge if we consider another sort of examples.

Let’s suppose Oscar states “I believe that I understand that fish breathe in water because I also believe that fish’ gills are able to extract oxygen from water”. It seem that this case fails to be a good piece of self-knowledge in slow-switching cases, since in Twin-

Earth water is not composed by H_2O . In this case, Oscar is mistaken about his own understandings²⁶. His thoughts about his own understanding could be corrected by an expert in Twin Earths, but this would certainly require the acquisition of further information about one's environment. And this fact, in Boghossian's argument, leads to incompatibilism. But again, the unacceptable point of the incompatibilist argument seems to be that Oscar is required to always be able to know his own thoughts in a direct and non-empirical way. However, appealing to this condition in order to deal with privileged self-knowledge is neither required nor acceptable. Lots of times we are aware of our thinkings in an indirect way, because someone has called our attention to some aspect of our behaviour or because we ourselves have engaged in some kind of self-analysis.

Therefore, the thought experiment in question –where the maintenance of the externalism is the supposed reason to the failure of one stance of privileged self-knowledge– could be used to sustain incompatibilism only in case we were assuming a very specific conception of self-knowledge: wherever it is part of this realm it must be entirely knowable a priori and directly. It seems that nowadays we have a lot of data favouring the denial of this conception²⁷.

2.3. Memory

²⁶ That fish breathe in water and that fish breathe in twater are both true. In this sense, Oscar's belief that fish breathe in water will be true, whatever concept he employs, *water* or *twater*. That fish extract oxygen from water is true in Earth but false in Twin Earth. Oscar would have a true belief in case he employs the concept *water* but false in case he employs the concept *twater*. Because of this, his reasoning about his own understanding is misled. Once Oscar has both concepts, he cannot make the link between his belief that fish breathe in water and his belief that fish extract oxygen from water without knowing which one he is using.

²⁷ This kind of compatibilist answer doesn't constitute an approach on self-knowledge, neither this is the aim of this paper. However, it suggests an important condition to an approach that wants to maintain externalism at the same time: it should be able to accommodate both methods of acquisition of self-knowledge.

The second line of argument favouring incompatibilism attempts to show that once slow-switching takes place, and externalism is considered to be true, there is no way of making sense of the memory of one's own thoughts. Several authors have found this argument also in Boghossian's comments and can be restructured as follows:

- (1) If S does not forget anything, then whatever S knows at time t1, S knows at time t2.
- (2) In the cases at hand S does not forget anything.
- (3) S does not know that p at time t2.
- (4) So S does not know that p at time t1²⁸.

²⁸ Cf. Burge, "Memory...", cit., p. 356. Burge considers this the only interpretation of Boghossian's argument. He says: "Much of the literature on this subject deals with problems that arise from the assumption that we need to *identify* the content of our thoughts in such a way as to be able to rule our relevant alternatives to what the content might be. Boghossian, unlike many of those who write on this subject, seems to recognize that this assumption is not acceptable on my view. One's relation to one's content, when one is non-empirically self-attributing in the reflexive, that-clause way is not analogous to a perceptual, identification relation to which alternatives would be relevant. In present tense self-attributions of the relevant kind, alternatives are irrelevant. Boghossian's strategy is to consider cases of memory and argue that these cases reflect badly on my view about the present tense cases" (Burge, *ibid.*, p. 355). However, if one reads through Boghossian's comments it seems that his argument does offer a double interpretation, especially because in the first formulation, Boghossian dedicates a good space to differentiate cases where relevant alternatives matter and where the problem is only about logical possibilities. As far as both formulations have received equal importance, I am considering both as valid.

Brueckner ("Externalism and Memory" in Ludlow, P. & Martin, N. (eds.), *Externalism and Self-knowledge*, cit., pp. 319-331.) does not only agree with Burge's reading, but he also thinks that Boghossian's argument is directed to Burge's account of basic self-knowledge. Nevertheless, Brueckner will conclude that "no Boghossian-style argument succeeds in refuting Burge's account of basic self-knowledge" and adds to it: "the covariation strategies are untouched as well" (Brueckner, *ibid.*, cit., p. 330). I have considered that Boghossian's argument is much more general than applied just to basic self-knowledge.

Let's remember Oscar, the subject that undertakes the switches. Let's suppose that just after one set of twin-earthian concepts has been displaced by a set of earthian ones, someone were to ask Oscar whether he had been recently thinking thoughts involving an arthritis-like concept distinct from arthritis. He would presumably say 'no'²⁹. But the fact is that, according to externalism, Oscar does entertain thoughts which involve twin concepts. The question that arises here is about how to explain this sort of "poor" ability to know past thinking. Once this does not seem to correspond to a bad capacity of remembering them properly, Boghossian suggests that in those cases, Oscar in fact never knew them.

Boghossian claims that although Burge is able to say that at t1 Oscar knows what he is thinking at that moment, he must accept that at t2 Oscar would fail to know what he was thinking at t1, exactly because the self-verifying character of basic self-knowledge applies only to current thoughts. Boghossian³⁰ understands that:

By Burge's criteria (...) [S] counts as having direct and authoritative knowledge at t1 of what he is thinking at that time. But it is quite clear that tomorrow he won't know what he thought at t1. No self-verifying judgment concerning his thought at t1 will be available to him then".

Again, it seems that Oscar would have to discover features of his environments in order to know what he himself thought in t2, exactly because such a thought would refer to the thought entertained in t1, and so, would not be self-verifying.

Burge replies to this formulation of the incompatibilist challenge by denying (P3). His fundamental idea is that "memory is fixed by the content of the thinking that it recalls"³¹, an idea developed as follows:

Memory need not be *about* a past event or content at all. It can simply link the past thought to the present, by preserving it. Such cases involve a particular type and function of memory

²⁹ Boghossian, "Content and self-knowledge", cit., p. 160.

³⁰ *Ibid.*, p. 171.

³¹ Cf. Burge, "Memory and self-knowledge", cit., p. 357.

–preservative memory– which preserves propositional contents and attitudes toward them, rather than *referring* to objects, attitudes, contents, images, or events³².

In the memory case, the content and referent of the remembered material is not distinct from that of the antecedent thought content, which in ordinary that-clause-type self-attributions is both thought and referred to³³.

The crucial point to Burge's defence is the differentiation between "preservative memory" and "memory by discrimination", and the insistence that the first is also essential to understanding such a phenomenon. This difference corresponds to the double interpretation of the question about whether an individual "knows what he was thinking yesterday"³⁴. If S relies upon memory to identify a past object or event –including a past thought– S will be subject to error; nevertheless, if S thought yesterday that twaluminum is beside him, he is in a position, relying on preservative memory, to remember what he thought then³⁵. The difference between both situations lies in the difference between a content being fixed in a past thought that is recalled in the present and the other situation where the present thought refers to a past one. The latter idea is not what Burge means by the function of preservative memory.

Following this reasoning, (P3) is clearly false on Burge's account. Just in case one had discrimination in mind, one could infer that in slow switching cases S does not know what he was thinking yesterday because he is unable to discriminate between two seemingly relevant possibilities³⁶. But, in preservative knowledge S knows that p at time t2. As Burge puts it,

[p]reservative memory normally retains the content and attitude commitments of earlier thinkings, through causal connections to the past thinkings³⁷.

³² *Ibidem*.

³³ *Ibid.*, p. 359.

³⁴ *Ibid.*, p. 362.

³⁵ *Ibid.*, p. 367.

³⁶ *Ibid.*, p. 362.

³⁷ *Ibid.*, p. 357.

Another important point insisted upon by Burge is how the second premise must be defended –(2): in the cases presented S does not forget anything. Boghossian supposes that when switches take place, one set of concepts is displaced by the other one, while Burge does not support such extravagance³⁸. To Burge, by no means S forgets one set of concepts when they are replaced by their counterparts. His proposal is to think about a scenario where “the individual has, without realising it, both the original concept and a new concept after slow-switching”³⁹. In this sense, premise (2) is completely defended by Burge, because the original beliefs are not forgotten, even if the subject can fail to access them in certain circumstances. Burge⁴⁰ stresses that:

Displacement was never part of the switching cases, at least in my understanding of them. Cohabitation was always the assumed case. I did not and do not consider the displacement model (as a general model for switching cases) a plausible account.

Burge offers another criticism over Boghossian’s argument in the sense that if displacement is behind such an argument, it seems that premise (2) is mistaken. He says

if one loses a concept when it is replaced by a new one, and for that reason one has no access to beliefs one once had, one may lose knowledge one once had⁴¹.

In this sense, the argument seems to fail in Boghossian’s very framework⁴².

³⁸ Actually Boghossian acknowledges both options of reading slow-switching cases but he finally endorses the “displacement” model.

³⁹ Burge, “Memory and self-knowledge”, cit., p. 368.

⁴⁰ *Ibid.*, pp. 364-365.

⁴¹ *Ibid.*, p. 369.

⁴² An alternative response to incompatibilism could be inspired by Ludlow’s comments. (“Social Externalism, Self-Knowledge, and Memory” (1995b) in Ludlow, P. & Martin, N. (eds.), *Externalism and Self-knowledge*, cit., pp. 307-310) He insists on the falsity of (P1). Ludlow claims that “Boghossian is correct in asserting that I do not know at t2 what I knew at t1, but he is incorrect in supposing that “the only explanation” for this is that I “never knew” my thoughts in the first place” (Ludlow, *ibid.*, p. 310). According to

According to this, the argument is unlikely sustained: on the one hand, if we insist that there is something like preservative memory, we should deny (P3); on the other hand, if we insist on the very Boghossian's view, that Oscar has his mental content

Ludlow, "It is entirely consistent with the social externalist view of memory that I forgot nothing, but that the contents of my memories have nonetheless shifted. Indeed, this is not only possible according to social externalism, but given the prevalence of slow switching it should be a rather common state of affairs" (Ludlow, *ibid.*, p. 310).

Ludlow's position has some serious problems. He claims that Boghossian's argument depends on an individualistic assumption about the nature of memory. According to Ludlow, the contents of our memories are subject to the same external conditions as every mental content is, and he understands by this that those external conditions must be the current ones. One of his serious problems is that he has a misconception of what is the most appropriate externalist account of memory.

Ludlow maintains that social externalism "is bound to say that the content of a memory is fixed at the time recollection takes place" (Ludlow, *ibid.*, p. 308). Otherwise, he says, one must accept that those contents are totally inert to all environment changes, and this seems to be contrary to externalism. Ludlow sees a problem in considering memory content somehow as "frozen up" to some later moment of recollection coexisting with the thesis that such contents are fixed by our social environment (Ludlow, *ibid.*, p. 309).

However, in the case of mental content of memories, there is no problem at all in accepting that their individuation factors held in the past. After all, memory is about the past. It is about recalling a past thought, with its past content, no matter what the current situation is. There is nothing problematic in being externalist and accepting it. The point is that externalism is not committed to the idea that mental contents are fixed by current external factors, but instead that such contents are individuated by external factors. And the history of this dependence relation matters here.

It seems that Ludlow's solution, in order to solve the incompatibilist challenge, turns the phenomena of memory into a completely empty and absurd faculty. For memory is about to recall the same thoughts one had entertained in some circumstance in the past. Once content of memory is taken to be individuated by current factors, memory no longer can do what it was supposed to do. In this sense, the immediate conclusion would be that one can seldom remember the thought one had earlier. And this is also quite unacceptable.

Ludlow doesn't seem to have many resources to avoid such criticisms. And it seems clear that his mistake is to suppose that externalism must take memory as he describes. What he conceives as memory cannot, after all, be classified as such.

replaced according to each world where he is located, and because of that, he doesn't know if he thinks about water or twater, (P2) is in danger. This kind of dilemma could be used to refuse such an argument as constituting a real risk to compatibilism.

3. *Thought experiments and compatibilism: Tyler Burge and Donald Davidson*

So far we have seen two incompatibilist instances suggested by the thought experiment of switching cases. However, we could go back and question about the reason why such contexts have received a privileged role within the philosophical debate. It is not obvious how this kind of thought experiment has anything to do with testing the idea that "if externalism obtains, then privileged self-knowledge doesn't", unless one has already supposed the problem to be the following:

To understand how we could know some of our mental events in a direct, non-empirical manner, when those events depend for their identities on our relations to the environment⁴³.

It is by translating the incompatibilist risks in those terms that it becomes clear how slow switching cases match this puzzling intuition.

Burge indicates that even in an extreme scenario, where one's own thoughts are individuated by external factors which are unknown to the subject of the experiment, such a subject is still able to know some of her thoughts in a privileged way. In this sense, Burge highlights compatibilism by reasoning about a scenario where privileged self-knowledge is not undermined by a failure in one's knowledge of one's environment.

Those conditions are, in a sense, very similar to the conditions that hold in the Cartesian demon thought experiment, where one could have direct and non-empirical self-knowledge while doubting completely the existence of a physical world. Skepticism is not in question here, but it is important to notice that part of Burge's strategy lies in insisting that the inference from the

⁴³ Cf. Burge, "Individualism and self-knowledge", cit., p. 650. It was actually Burge who first indicated this sort of puzzle.

Cartesian account on self-knowledge to individualism is misleading⁴⁴. In fact, Burge makes it clear that part of his aims is to sustain a “restricted Cartesian conception of self-knowledge”⁴⁵.

If this is so, if part of the compatibilist task lies in deconstructing the connection between individualism and this kind of approach on self-knowledge, maybe we could dispense with the use of a thought experiment⁴⁶. It would be enough to consider what John Heil suggests:

If the contents of one’s thoughts were determined entirely by the state of one’s brain, why should this fact alone make our access to them any less indirect or difficult?⁴⁷

It seems clear that if one had a “Cartesianism’s caricature” in mind, internalism and total access to one’s own mind were to be blended in one and the same position. But if one departs from the question about the nature and the individuation of mental contents, internalism is not equal to total access to one’s own mind. It is instead a position which defends that one’s mental states are to be individuated by internal factors to the head, such as brain states. As Heil⁴⁸ indicates, there is no clear point in saying that just externalist theories of contents could motivate doubts about the possibility of privileged access.

So, it seems that the inference from privileged self-knowledge to individualism could be easily undermined if we realised that internalism can be much wider than Cartesianism. However, there are other questions involved in the compatibilist enterprise. As Burge himself indicates:

⁴⁴ *Ibid.*, pp. 651-652.

⁴⁵ *Ibid.*, p.649.

⁴⁶ Actually, Burge recognises it, saying that such an inference was already showed to fail by Arnauld’s comments on Descartes. Nevertheless, Burge suggests that undermining the Cartesian inference still leaves us with the puzzling sensation that there must be something wrong with externalism. I will insist on the step of rejecting this inference by showing that there are other individualist positions that should deal with the same problems externalism is accused of.

⁴⁷ Heil, J., “Privileged Access”, in *Mind*, vol. 97 1988, 386, p. 247.

⁴⁸ *Ibidem*.

It is one thing to point out gaps in inferences from self-knowledge to individualism. It is another to rid oneself of the feeling that there is a puzzle here⁴⁹

And Burge's slow switching cases deals with another important question: the dependence between kinds of knowledge.

It seems that, for Burge, answering a question such as:

Why is our having non-empirical knowledge of our thoughts not impugned by the fact that such thoughts are individuated through relations to an environment that we know only empirically?⁵⁰

Involves arguing in favor of the independence of kinds of knowledge: self-knowledge and world's knowledge. Taking slow switching cases to be a good context of discussion seems to localize Burge's position very close to the skeptic's, because instead of considering such cases as abnormal ones, Burge prefers to state that self-knowledge is left untouched while one can be completely ignorant about one's own environment⁵¹.

Davidson, who agrees with Burge in defending compatibilism, "[does] not consider Burge's thought experiments as persuasive as he does"⁵², maybe because he defends that self-knowledge and knowledge of the world are interdependent. Furthermore, they are

⁴⁹ Cf. Burge, "Individualism and self-knowledge", cit., p. 652.

⁵⁰ *Ibid.*, pp. 652-653.

⁵¹ I have, however, exaggerated Burge's position here. The subject's ignorance is localized. Actually, the only statement Burge commits himself to is that there is not an easy answer to the skeptic through externalism. However, once Burge's externalism demands that those proper connections between mind and world must have occurred in order to one's possession of thoughts, he seems to avoid some skeptical worries. In fact, he would not accept general skeptical scenarios so easily; he would first ask the skeptic to explain how the deluded individual has acquired his concepts; and second, if the answer was that the demon has induced him, Burge would argue that the demon would probably have had connections with the world. Nevertheless, I will insist that slow switching cases seem to share some similarities with Cartesian thought experiments.

⁵² Davidson, D., "Reply to Burge", en *The Journal of Philosophy*, vol. 85, 1988, 11, p. 665.

also interdependent of knowledge of other minds. As Burge also does, Davidson accepts the following ideas:

That the contents of our thoughts are individuated in part on the basis of external factors of which the thinker may be ignorant, and that thinkers are authoritative with respect to the contents of their thoughts⁵³

But it seems that, for Davidson, the concern about how we can know our thoughts without knowing the world in advance must be dissolved instead of answered. The point is that we need world knowledge (as well as knowledge of other minds) in order to know our thoughts, but also the other way round. So, there is no question about priority here, nor a problem about world information being required as an enabling condition to knowledge about oneself. One's self-knowledge is also required in order to know the world.

Davidson states that the basic reason for him to hold compatibilism is that "what determines the contents of thoughts also determines what the thinker thinks the contents are"⁵⁴. In a sense, this totally coincides with Burge's position, yet it seems weaker than appealing to a range of self-verifying thoughts.

As already indicated, the most widely accepted compatibilist answer has been based on the fact that the second-order thought somehow involves the first-order thought, which is individuated externalistically. Such an element has been used to show that in fact "there is no *special* problem for the achievement of self-knowledge in the fact that my first-order thinking is subject to an externalist dependence thesis"⁵⁵, although this fact clearly does not explain by itself how it is that my second-order thought amounts to knowledge.

Davidson's compatibilism makes use of such an element. However it is important to recognize that not only his theses about radical interpretation but also the one about the interdependence

⁵³ *Ibid.*, p. 664.

⁵⁴ *Ibidem.*

⁵⁵ Cf. Davies, M., "Externalism and Armchair Knowledge" (2000) in Boghossian, P. & Peacocke, C. (eds.), *New Essays on the A Priori*, Oxford, cit., p. 391.

between the three kinds of knowledge play a decisive role both in Davidson's externalism and in Davidson's compatibilism⁵⁶. As Heil⁵⁷ points out, Davidson's compatibilism indicates that the problem lies not in how externalism deals with privileged self-knowledge, but in a problematic "picture of mind", that needs to be solved. It is a picture where

beliefs about the contents of one's mental states are taken to be based on inward glimpses of those states or on the grasping of particular entities (contents, perhaps, or propositions, or sentences in mentalese)⁵⁸

Davidson recommends that we abandon the notion that knowledge of mental contents requires our inwardly perceiving in such a way. Once we do so, we remove at least one of the reasons for assuming that externalism undermines privileged access.

This picture of mind is not maintained by Burge either, but there are some important remaining differences between both compatibilisms which seem to refer back to the dependence or independence between kinds of knowledge. Once Burge doesn't see a problem with stating independency, slow-switching cases gain more interest to him than to someone like Davidson, who doesn't see it as a good solution.

Considering the thought experiment as such, Burge seems to provide a consistent compatibilist answer when he maintains his externalist view while appealing to the characteristics of basic self-knowledge. If the question was about the possibility of finding privileged self-knowledge in an externalist framework by offering a range of cases where the answer is positive, Burge reaches a reasonable compatibilist solution.

However, I have suggested that compatibilism could be maintained without giving an answer to such cases. It also seems to

⁵⁶ Cf. Davidson, D., 'Radical Interpretation' (1973), in *Inquiries into Truth and Interpretation*, Oxford, Clarendon Press, 1984, pp. 125-139; Davidson, D., "Three Varieties of Knowledge" (1991), in *Subjective, Intersubjective, Objective*, Oxford, Clarendon Press, 2001, pp. 205-220.

⁵⁷ Cf. Heil, "Privileged ...", cit., p. 247.

⁵⁸ *Ibidem*.

be dispensable to insist on Cartesian intuitions in order to talk about privileged self-knowledge. Burge is sympathetic to a restricted Cartesian approach to self-knowledge. But, if by “restricted” Burge means that just a part of our self-knowledge is acquired in a direct and non-empirical way, there is no need at all to insist on the label “Cartesian”. A restricted thesis does not seem to be a Cartesian thesis anymore, especially considering that the second-order beliefs partly inherit their content from externalistically individuated beliefs.

This suggests that there are two available paths for the compatibilist to deal with switching cases: to search for an answer to the proposed challenge while maintaining its initial conditions, as Burge seems to do, or merely solving it, as it seems Davidson does. However, neither of those paths seems to be enough to the establishment of compatibilism, once there is a second context of discussion. A context that would remain intact, even if all the possible problems arisen with the mental experiment are solved: the *reductio ad absurdum* arguments.

4. *Reductio ad absurdum of compatibilism*

The second context of discussion where compatibilism has been tested was initially indicated by McKinsey⁵⁹, but has acquired several formulations, such as Boghossian’s:

Let’s suppose that Oscar [...] is a compatibilist. I claim that Oscar is in a position to argue, purely a priori, as follows:

[P1] If I have the concept *water*, then water exists.

[P2] I have the concept *water*.

Therefore,

⁵⁹ Cf. McKinsey, M., “Anti-Individualism and Privileged Access”, in *Analysis*, vol. 51, 1991.

[C3] Water exists⁶⁰.

According to Boghossian, (P1) is reached non-empirically by philosophical arguments that sustain externalism while (P2) constitutes Oscar's privileged self-knowledge. Therefore, (C3) could be concluded also by a non-empirical way. And this is the element used against compatibilism: to know a fact of the world, such as the fact that water exists, by a non-empirical manner would be something absurd.

There are several available strategies in order to avoid the alleged incompatibilist result. We could enumerate them as follows: (1) To refuse one of the premises; (2) To defend that the conclusion is not indisputably unacceptable; and (3) to defend that the argument, although being a valid one, has problems that are revealed in terms of epistemic warrants of its elements and how they are related to each other.

The second strategy is emblematically defended by Sarah Sawyer⁶¹, who argues that inferences from introspective knowledge to empirical knowledge are not to be seen as intrinsically unacceptable⁶². To consider them as such would constitute a dogma, if our starting point is already externalist. There is nothing epistemically wrong with the argument⁶³. Yet it would be necessary to understand that, for an externalist, to know the world through self-knowledge is not too much to ask, because the concepts of this realm are not themselves unconnected with the world. In order to acquire a concept, a causal connection between the world and my mind is necessary⁶⁴.

⁶⁰ Boghossian, P., "What the Externalist Can Know 'A Priori'", *Philosophical Issues*, vol. 9, 1998, p. 202.

⁶¹ Cf. Sawyer, S., "Privileged Access to the World", *Australasian Journal of Philosophy*, vol. 76, 1998, 4, p. 528. Warfield ("A Priori Knowledge of the World: knowing the world by knowing our minds", *Philosophical Studies*, vol. 92 (1998), pp. 127-147) has a similar strategy.

⁶² *Ibidem*.

⁶³ Sawyer, S., "Externalism, Apriority and Transmission of Warrant" (2006) in Marvan, T. (ed.), *What Determines Content? The Internalism / Externalism Dispute*, Cambridge, Cambridge Scholars Press, 2006.

⁶⁴ Cf. Sawyer, "Privileged Access to" ..., cit., p.528

The third strategy has gained a very interesting dimension and, in fact, could be developed under different sub-strategies⁶⁵. Wright's and Davies's analyses represent important strategies within this group. Although they maintain important differences between their approaches, both of them indicate that in the argument in question the epistemic warrant of the premises is not transferred to the conclusion. According to Wright⁶⁶, despite the above argument being a valid one, it is not a cogent argument, because the premise's justification seems to require prior epistemic warrant of the conclusion. In this way, the argument would lack the distinguishing feature of leading someone to learn the truth by the justification of the premises, which is the fundamental characteristic of a cogent argument⁶⁷.

Although the above strategies establish important paths in order to deal with the incompatibilist challenge, it seems that the first one would deserve more of our attention because it concerns the very commitments of a compatibilist. On the one hand, it analyzes what externalism would enable us to know, and on the other hand, what kind of self-knowledge we would have. If our fundamental matter was the incompatibilist discussion, to maintain the argument is a serious mistake, if it is built upon misleading premises.

In the forthcoming lines I shall defend that the argument depends on a misleading conception of externalism. If, as Sawyer herself points out,

⁶⁵ Cf. Sawyer (*ibid.*) for an overview of the available options. In such a context, the incompatibilist argument is usually taken as a particular instance of a type-argument, of which Moore's proof of the external world is another case.

⁶⁶ Cf. Wright, C., "Some Reflections on the Acquisition of Warrant by Inference" (2003) in Nuccetelli, S. (ed.), *New Essays on Semantic Externalism and Self-Knowledge*, Cambridge Mass, MIT Press, 2003.

⁶⁷ *Ibid.*, p. 57.

The example is obviously problematic, since no reasonable form of externalism would support the linking conditional stated in [P1]⁶⁸

It seems that the urgency lies in deconstructing the argument in that direction.

Boghossian⁶⁹ anticipates two possible ways of rejecting P1: 1. that water would not be required for the acquisition of the concept of water; or 2. that water is required for the acquisition of the concept of water, but this fact could not be known *a priori*. Boghossian argues that such possibilities are easily ruled out, therefore giving rise to incompatibilism. However, Burge⁷⁰ argues against it, insisting on (1) while Goldberg⁷¹ insists on (2). Burge claims that:

Despite its extreme schematic character, this principle [P1]⁷² - or any instance of it- is false. As I pointed out in "Other Bodies"; water need not exist in an individual's environment in order for the individual to think that water is such and such⁷³.

Burge suggests that

if one is sufficiently precise, one could introduce a 'natural kind' notion, like water without having had any causal contact with instances of it⁷⁴

⁶⁸ Sawyer, "Externalism, Apriority and ...", cit., p. 6. [P1] replaces W2 for the sake of text's coherence. In the original text, W2 is the following premise: "If I think that water is wet, then there is water in my environment"

⁶⁹ Boghossian, "What the Externalist...", cit., pp. 197-211.

⁷⁰ Burge, T., "Replies from Tyler Burge" (2003) in Frápolli, M.J. & Romero, E. (eds.) *Meaning, Basic Self-knowledge and Mind*, cit., pp. 243-296.

⁷¹ Goldberg, Sanford, "On Our Alleged A Priori Knowledge that Water Exists", *Analysis*, vol. 63, 2003, 277, pp. 38-41.

⁷² Burge refers to the following principle:
"WaterDeep Necessarily, for all x, if x is thinking that water is wet then x is (or has been) embedded in such-and-such ways in an environment that contains samples of water" *Ibid.*, 262.

⁷³ Cf. Burge, T., "Replies from Tyler Burge" (2003) in Frápolli, M.J. & Romero, E. (eds.) *Meaning, Basic Self-knowledge and Mind*, cit., p. 262.

⁷⁴ Cf. Burge, T., "Other Bodies" (1982) in *Foundations of Mind*, Oxford, Oxford University Press, 2007, p. 98.

He reminds us that some sciences such as chemistry have indeed anticipated some natural kinds before their discovery “in nature”. Externalism doesn’t need to deny such a fact.

More than that, Burge adds that an individual or a community could have been mistakenly thinking that there was something such as water. And the point is that if this mistake was discovered, the concept would not be completely emptied⁷⁵. Burge insists that⁷⁶:

As I previously indicated, I think that Adam’s having attitudes whose contents involve the notion of water does not entail the existence of water. If by some wild communal illusion, no one had ever really seen a relevant liquid in the lakes and rivers, or had drunk such a liquid, there might still be enough in the community’s talk to distinguish the notion of water from that of twater and from other candidate notions. We would still have our chemical analyses, despite the illusoriness of their object. [...] I think that Adam’s having the relevant attitudes probably does not entail the existence of other speakers. *Prima facie*, at least, it would seem that if he did interact with water and held a few elementary true beliefs about it, we would have enough to explain how he acquired the notion of water. What seems incredible is to suppose that Adam, in his relative ignorance and indifference about the nature of water, holds beliefs whose contents involve the notion, even though neither water nor communal cohorts exist.

Goldberg⁷⁷, on the other hand, insists that the problem with P1 lies elsewhere. Regarding the argument as formulated above, water is indeed a necessary condition for the possession of the concept of water, but such a fact could not be known a priori. He claims that⁷⁸:

The upshot is that McKinsey-style arguments, which would have us conclude [...] that I can know a priori that e.g. water exists, fail, for assuming that all statements expressing

⁷⁵ *Ibid.*, p. 97.

⁷⁶ *Ibid.*, p. 98.

⁷⁷ Goldberg, “On Our Alleged A Priori Knowledge that Water Exists”, in *Analysis*, vol. 63, 2003, pp. 38-41.

⁷⁸ *Ibid.*, pp. 40-41.

metaphysical dependencies between their designata are knowable a priori (...)

Precisely not, since the metaphysical dependence of WATER on the existence of water (H₂O) itself depends on the identification of water with H₂O.

Although the latter may be questionable (that the metaphysical dependence of *water* on the existence of water (H₂O) itself depends on the identification of water with H₂O) it seems fair to accept that the question about whether water is necessary for the acquisition of the concept of *water* is a matter of empirical knowledge. In that sense, this indicates an alternative route to reject P1. Nevertheless, such a route seemingly has a very narrow application. It could only be applied to the kind of externalism deduced from Putnam's works⁷⁹ and which is the base of the incompatibilist argument proposed by Boghossian⁸⁰.

In Putnam's context⁸¹ (1975), if the external trait of Oscar's mental states is explained by the fact that water has caused such a thought, it seems that P1 would be available to an externalist⁸². If Oscar has the concept of water, and he is an externalist, the fact that water exists would be available for him. If it is discovered that he was wrong, that in fact water doesn't exist, what Oscar had was not a concept but a pseudo-concept instead. It is here where Goldberg's criticism has an application. An externalist would be able to reach P1 because, besides knowing the philosophical arguments that have led him to P1, he had knowledge of the world, in this case, about the constitution of water. P1 could be the Putnamian lesson in 1975, and so it could be attacked following Goldberg's criticism, but it should not be mixed with the several available externalist positions.

⁷⁹ Putnam, "The Meaning of...", cit., pp. 215-271.

⁸⁰ Boghossian, "What the Externalist...", cit., pp. 197-211.

⁸¹ Putnam, "The Meaning of...", cit., pp. 215-271.

⁸² That Putnam is committed to (P1) seems to be connected with his conception of meaning as being composed by stereotypes plus reference. However, in what follows, I shall develop another line of argumentation. I will indicate one possible interpretation of Putnam's position that sees him as committed with the idea that reference points to the sufficient and necessary cause of one's thoughts.

This represents a third route to indicate that P1 has to deal with serious objections. P1 is based on an externalist position which is neither the unique nor the prevailing one. An externalist does not need to sustain that the external trait of my concepts is due to the supposed fact that the related objects of my concepts have caused them. At least, not in the atomist way as it seems to be assumed by the incompatibilist argument.

An externalist position emphasizes that the mind is constituted by the external to our skin because we interact with our world and with our community. Some positions explain such interaction appealing to causality, explaining, for instance, that our mind is constituted by what is external to our skin through causal relations between oneself, one's fellows and one's world. Others prefer to explain such interaction by appealing to our linguistic abilities and to the notion of objectivity, explaining that our mind is constituted by what is external to our skin because the base of our mental realm is constituted by a reasonable range of knowledge. However, it seems that just a few positions would sustain that it is possible to deduce, from each of our concepts, a correspondent object to which we could refer in order to explain the history of the acquisition of that concept. In an externalist framework, a mental holism seems to have much more space than an atomism. An atomism seems to require, in fact, that some of our mental contents need to be identified in an internalist manner.

Let's consider an atomist position in which each of our concepts should correspond to an item of the world. Such a position is clearly problematic, once we have concepts "without correspondents" in the world, such as the well-known example of the unicorn. How should those concepts be individuated? If we follow the atomist line of reasoning, they might be individuated by an internalist manner. If there is no such a correspondent in the world, we might explain them as being pseudo-concepts or in terms of something internal to our heads. In this way, when an atomist considers himself able to explain the external character of some of our concepts, he only achieves this by maintaining another broad group of concepts individuated by an internalist way.

However, the aim of several externalist positions is to sustain that at least part of all our mental contents are constituted by an

external manner. The very idea of internalistically individuated contents –the narrow contents– has been under attack under the accusation of being untenable notions. This has also been an important criticism directed to Putnam’s position in 1975⁸³, when his externalism was sustained upon the price of the necessity of narrow contents.

In his way, although Goldberg’s criticism seems to be a good one, it has its own scope diminished because he contemplates a condition only sustained by a specific and controversial kind of externalism. Because of that, it seems that the most reasonable thing to do would be to insist that P1 is not an externalist consequence, by the reasons indicated above as well by Burge’s reasoning. In doing so, the argument in question here could not serve as the basis of an incompatibilist attack.

5. Conclusion

Despite the variety of questions and arguments treated in the text, I have defended a general thesis in the following terms: the incompatibilist challenges introduced in the literature under the form of slow-switching cases and *reductio ad absurdum* arguments represent a real objection to compatibilism only under very specific conditions: if we assume a specific account of self-knowledge in the first case and if we assume a specific approach of externalism in the second case. When we move on to other approaches, the incompatibilist risk is solved.

In the first part, I have treated the slow-switching cases such as exposed by Burge⁸⁴. I’ve discussed two incompatibilist arguments based on Boghossian’s comments⁸⁵ (1989) and some of the ways they could be answered.

The first interpretation of Boghossian’s comments has pointed out that the switched subject could not have knowledge of her own thoughts once she was unable to discriminate between water and twater thoughts. The second argument has driven a criticism over compatibilism appealing to questions about memory. Taking

⁸³ Putnam, “The Meaning of...”, cit.

⁸⁴ Burge, “Individualism and self-knowledge...”, cit., pp. 649-663.

⁸⁵ Boghossian, “Content and Self-knowledge”, cit., pp. 149-173.

Burge's framework about self-knowledge as the starting point, the argument has indicated that even though the subject of the experiment could know her own thoughts in a privileged manner at the moment she was thinking them, she would be unable to remember them later on.

Considering Burge's comments on that question, the latter argument would lead us to a kind of impasse: on the one hand, if we accept that there is something like preservative memory, we would have to disregard (P3). On the other hand, if we maintain (P3), we would have to abandon (P2). Therefore, I have defended that such an argument could not provide a basis for an incompatibilist position.

Regarding the former argument, I have defended that it could only be used as a support for incompatibilism if we insisted on the following view about self-knowledge: all that deserves the label of self-knowledge might be potentially available to be known in a direct and non-empirical manner. However, I've argued that such a vision about self-knowledge would be so problematic as to sustain that there is no parcel of privileged self-knowledge. A fair account of self-knowledge should give rise to the privileged kind of acquisition as well as to the indirect and empirical method by means of which we know part of our minds.

In the intermediate part, I have raised the question about the role the thought experiment occupies in the attempt to defend compatibilism. I've discussed two compatibilist frameworks, Burge's and Davidson's. While Burge proposes the experiment, offering a compatibilist answer to it, Davidson would instead tend to solve it. I have suggested that Davidson is also able to provide a compatibilist framework following a different route, one that dispenses with Burge's commitments, such as the reference to basic self-knowledge and his supposed commitment to the thesis of independence between kinds of knowledge. I have also referred to the idea that the external character of the second-order thoughts is due to the fact that they somehow embed the first-order thoughts, which are externalistically individuated. In this way, there would not be a special problem about how to explain the privileged acquisition of self-knowledge while externalism is in place. The

vertigo of puzzle, however, should be cured together with the dissolution of the Cartesian approach to self-knowledge.

The last part of the text has treated the second group of incompatibilist challenges, the *reductio ad absurdum* argument, as exposed by Boghossian⁸⁶ (1998). I have defended that if such an argument was based on a misleading conception of externalism, it should be rejected as a good support for incompatibilism. Taking into account Burge's⁸⁷ and Goldberg's⁸⁸ reasons to reject P1, it was defended that Putnam's externalism⁸⁹ would be the only instance where P1 might have space. That is, disregarding Goldberg's criticism⁹⁰, the only kind of externalism that could have P1 as a consequence of its theses would be an externalism of Putnam's type. It was argued that externalism, in general, neither needs nor is committed to the implication involved in P1. To sustain that our thoughts are identified in relation to external factors doesn't give us the right to infer the existence of a supposed correspondent of the mental content in the world. If there is in the world a referent of a particular concept, and if it had some importance in the acquisition of the thought in question, this acquisition would not be independent of the community or even of the very individual.

cborgoni@ugr.es
Universidad de Granada

⁸⁶ Boghossian, "What the Externalist...", cit., pp. 197-211.

⁸⁷ Burge, "Replies from Tyler...", cit., pp. 243-296.

⁸⁸ Goldberg, "On Our Alleged...", cit., pp. 38-41.

⁸⁹ Putnam, "The Meaning of...", cit., pp 215-271.

⁹⁰ Goldberg, "On Our Alleged...", cit., pp. 38-41.