

Semantic Externalism and *A Priori* Self-Knowledge¹

Jussi Haukioja

Abstract

The argument known as the 'McKinsey Recipe' tries to establish the incompatibility of semantic externalism (about natural kind concepts in particular) and *a priori* self-knowledge about thoughts and concepts by deriving from the conjunction of these theses an absurd conclusion, such as that we could know *a priori* that water exists. One reply to this argument is to distinguish two different readings of 'natural kind concept': (i) a concept which *in fact* denotes a natural kind, and (ii) a concept which *aims* to denote a natural kind. Paul Boghossian has argued, using a *Dry Earth* scenario, that this response fails, claiming that the externalist cannot make sense of a concept aiming, but failing, to denote a natural kind. In this paper I argue that Boghossian's argument is flawed. Borrowing machinery from two-dimensional semantics, using the notion of 'considering a possible world as actual', I claim that we can give a determinate answer to Boghossian's question: which concept would 'water' express on Dry Earth?

I

Semantic externalism, as understood in this paper, is the view that the propositional content of an intentional state of an agent is, at least in some cases, partly constituted by factors external to the agent. During the last fifteen years different kinds of arguments have been presented for *incompatibilism*, the view that such externalism is incompatible with certain wide-spread views about self-knowledge. It is usually thought that, when I am thinking that water is wet, I can know, *a priori*,² that I am thinking that water is wet. That is, we can get introspective knowledge about the contents of our own thoughts. Both externalism and the view that *a priori* self-knowledge is possible are so wide-spread that it would be quite remarkable if they were to turn out to rule out one another.

In this paper I will focus on one argument for incompatibilism, the so-called 'McKinsey recipe'.³ In this argument, the incompatibilist claims to find a proposition, P, which we clearly

cannot know *a priori*, but which could be derived from the conjunction of externalism and *a priori* self-knowledge (McKinsey 1991, pp. 12-16). When we are dealing with natural kinds, the most promising candidate for P seems to be the proposition that *instances of the natural kind exist*, giving us the following argument:

(A) I can know *a priori* that: if I have the concept WATER⁴, water exists.

(B) I can know *a priori* that I have the concept WATER.

therefore,

(C) I can know *a priori* that water exists.

The argument is supposed to run as follows. Semantic externalism implies (A). Our concept WATER, according to externalism, is the concept of water (and not, for example, twater), precisely because water (and not, for example, twater), exists on Earth. When Oscar uses the term 'water', he refers to water because he has interacted causally with instances of water, and this would not be possible if water did not exist. On the other hand, the concept Toscar, a denizen of Twin Earth, expresses with the term 'water' is not the concept of water, but the concept of twater, precisely because the instances Toscar has interacted with are instances of twater.⁵ The apriority in (A) follows from the fact that externalism is supposed to be a *conceptual* truth, not an empirical one. Hence, if the claim 'if I have the concept WATER, water exists' follows from externalism, and externalism is true, the claim can be known *a priori*. (B), on the other hand, follows directly from the claim that we can have *a priori* knowledge about the contents of our own thoughts. If I can know *a priori* that I am thinking that water is wet, I can know *a priori* that I have the concept WATER. But (A) and (B) jointly imply (C)⁶, and (C) is unacceptable.⁷ Thus, either (A) or (B) must be false, and externalism and *a priori* self-knowledge have been shown to be incompatible.

II

Generalising away from the example of water, we get the following (using *A* for *a priori* knowability):

- (1) A(If I have natural kind concept K, instances of K exist)
 - (2) A(I have natural kind concept K)
- therefore,
- (3) A(Instances of K exist)

A natural response to this argument is to deny premiss (2) (McLaughlin & Tye 1998b, pp. 369-372). If we take a natural kind concept to be a concept which in fact denotes a natural kind, (1) does seem to follow from externalism. (2) would then claim that, in addition to knowing *a priori* that I have concept K, I could *also* know *a priori* that K is a natural kind concept. But the question of whether or not K denotes a natural kind seems to be an empirical one, and hence not answerable on *a priori* grounds. The most we can know *a priori* is that concept K 'aims' to denote a natural kind – that it somehow contains or involves the presumption that its instances share a naturalistic essence.⁸

Depending on what we take 'natural kind concept' to mean, we would thus get two different arguments:

Argument A:

- (1a) A(If I have concept K, and K denotes a natural kind, instances of K exist)
 - (2a) A(I have concept K, and K denotes a natural kind)
- therefore,
- (3) A(Instances of K exist)

or argument B:

- (1b) A(If I have concept K, and K aims to denote a natural kind, instances of K exist)
 - (2b) A(I have concept K, and K aims to denote a natural kind)
- therefore,
- (3) A(Instances of K exist)

Both arguments are valid, but neither is sound. In argument A, premiss (1a) follows from semantic externalism (and is presumably true), but as noted above, there is no reason to suppose that (2a) is true. Argument B, on the other hand, could be used to give an *a priori* 'proof' of the existence of, say, unicorns or phlogiston: if I have a concept which aims to denote the natural kind of unicorns, we can presumably know this *a priori* (2b). It would then follow from argument B that unicorns exist.⁹ The source of the problem is easy to see. Premiss (1b) is absurd, and it does not follow from any respectable form of externalism. From the mere fact that a concept *aims* to denote a natural kind it clearly does not follow that it *in fact* denotes anything. To summarise: with two readings of what it is to be a natural kind concept, we get two readings of the argument from (1) and (2) to (3). On one reading, premiss (2) is false (2a); on the other, premiss (1) is false (1b).

III

Paul Boghossian (1997) has claimed that this line of response fails. According to him, externalists cannot make sense of the possibility of a concept K aiming, but failing, to denote a natural kind. When a term such as 'phlogiston' has an empty extension, the externalist is, according to Boghossian, forced to say that it does not express any concept *at all*. If he is right, externalism would entail that no one can have the concept of phlogiston in the actual world, where phlogiston does not exist, and hence no one can really have thoughts about phlogiston, even though we may *think* we have such thoughts. Externalists would then be forced to say that the phlogiston theorists of the late eighteenth century did not really *believe* that phlogiston exists and that it is released in combustion, because there could not be the natural kind *concept* of phlogiston without the natural kind itself. They only *thought* they had the concept PHLOGISTON, and thought they believed that phlogiston is released in combustion. If

Boghossian is right, then the externalist should deny that introspection is sufficient for knowing the contents of our own thoughts.

If Boghossian really were right, and the above claims followed from externalism, this alone would, in my view, be reason enough to give up externalism, quite regardless of questions about *a priori* self-knowledge. Let us take a closer look at why he thinks the externalist is committed to such strange views about extensionless natural kind terms.

Boghossian bases his claim on a variation of the Twin Earth story (Boghossian 1997, pp. 170-174). Instead of assuming an Earth-like planet with XYZ, Boghossian asks us to imagine *Dry Earth*, where it *seems* to our counterparts that there is a clear, tasteless and odourless liquid flowing in their rivers, seas, swimming pools and so on. But in reality there is no such liquid. Dry Earthlings are under a collective mirage. There is not a single molecule of such liquid (water, twater, or any other kind) on Dry Earth. Despite the intuitive implausibility of Dry Earth, it is not qualitatively different from the situation in which Earthlings found themselves with respect to phlogiston in the late eighteenth century.

Boghossian now presents the following question to the externalist: which concept do Dry Earthlings express with their term 'water'? It seems obvious that their 'water' aims to denote a natural kind. According to externalism, the contents of concepts which aim to denote natural kinds are determined partly by which natural kind the concept possessors in fact happen to causally interact with. But Dry Earthlings have not, by assumption, interacted with *any* kind when it has seemed to them that they see, smell and taste a colourless, odourless and tasteless liquid. Boghossian does not claim that this directly commits the externalist to the claim that 'water' expresses no concept on Dry Earth, but he claims that in the end the externalist cannot give a satisfactory account of *which* concept, exactly, they do express with it.

To show this, Boghossian asks the externalist to specify the truth conditions of 'water is wet' on Dry Earth. According to him, we can give truth conditions to this sentence only if we

interpret 'water' as expressing a *compound* concept on Dry Earth – that is, only if we interpret it as having been constructed from other concepts. 'Water' would then express, roughly, the concept of 'that colourless, odourless and tasteless liquid which mainly fills the seas and waters etc.'. But, says Boghossian, the externalist cannot give this answer. According to externalism, the concept WATER, on Earth, expresses an *atomic* concept. The atomicity of a concept is something determined by our 'internal syntax', and we are assuming that Earthlings and Dry Earthlings are identical with respect to their inner workings, so the externalist cannot claim that 'water' would express a compound concept on Dry Earth and an atomic one on Earth. Hence, Boghossian concludes, the externalist cannot give the truth conditions of 'water is wet' on Dry Earth, and consequently cannot specify which concept the Dry Earthlings express with their tokens of 'water'.

IV

It seems to me that Boghossian completely leaves out of his consideration the most plausible externalist answer. It is true that externalists think that many natural kind terms, such as 'water', cannot be *defined* in other terms, and hence that the corresponding natural kind concepts are not constructed out of other concepts. It is one of the central conclusions of externalist arguments that, say, the term 'water' is not identical in meaning with any description of the kind 'the colourless, odourless and tasteless liquid which ...', because these expressions do not behave alike in modal contexts. So such descriptions are not *meaning-giving*. Nonetheless, descriptions of this kind *are* thought to function as *extension-fixing* or *reference-fixing* (in one of two ways; see section VI below). Some description of the kind 'the colourless, odourless and tasteless liquid which...' fixes the reference of 'water' and WATER – 'water' denotes the colourless, odourless and tasteless stuff which *in the actual world* mostly fills the seas and rivers and so on – the stuff that plays a certain role in our environment.

In setting up Twin Earth and Dry Earth, we are assuming that Toscar (a Twin Earthling) and Doscar (a Dry Earthling) are like Oscar (an Earthling) in their internal respects. Consequently, it appears to them that they are acquainted with a stuff which plays the watery role in their respective environments. If a plausible externalist story about natural kind term reference can be told, it will explain how, on the basis of Oscar's acquaintance with water (possibly augmented by social factors), it follows that a description will fix the extension of his term 'water' (again, in one of two ways; see section VI). But, since we are assuming Toscar and Doscar to be like Oscar in all internal (and social) respects, it will follow that the same story can be told about them as well. Thus, Doscar's 'water' will express the concept which is associated with the relevant reference-fixing description in exactly the same way as Oscar's 'water' is¹⁰, and for exactly the same reasons, but has an empty extension because nothing satisfies the description. For Doscar, the concept expressed by 'water' is just the concept of the familiar stuff which appears to play a certain role in his environment. It just happens that there *is* no such stuff.

V

A similar response to Boghossian has been given by McLaughlin & Tye (1998a) and Moya (1998). Moya goes on to claim that the sentence 'Water is wet', uttered on Dry Earth, is in fact false, but it 'would be true if there were in fact a substance, a natural kind, in the extension of "water" and this substance were wet.' (Moya 1998, p. 250) I think this is misleading at best. Since 'water' fails to refer on Dry Earth, Doscar's tokens of 'water' have an empty extension in *all* possible worlds (considered as counterfactual by Doscar, cf. below). Accordingly, if Doscar's utterance of 'Water is wet' is false, it will also be false when evaluated in worlds where a watery stuff *does* exist (and is available to Doscar's counterpart in that world etc.), because the 'stuff' in the extension of Doscar's 'water', (i.e, nothing) is not wet in those worlds.

However, I think the correct answer to Boghossian *can* be found along similar lines. But a more careful formulation is needed. The machinery of two-dimensional semantics¹¹ is, in my view, well suited to clarifying the situation here. (In fact, the Dry Earth story naturally suggests the use of the framework – when we are asking ourselves what the semantic value of Doscar's tokens of 'water' is on Dry Earth, what else is this but an invitation to consider Dry Earth as actual?) What we really should say about Doscar (and, perhaps, what Moya wanted to say) is that his tokens of 'Water is wet' would be true on (say) Earth, *considered as actual* by Doscar on Dry Earth. Because of the way Earth, Twin Earth and Dry Earth are set up, the same *primary* or *epistemic* intensions will be associated with Oscar's, Toscar's, and Doscar's uses of 'water'. That is, the referent of 'water' will, for each of them, depend in exactly the same way on how their world turns out to be. 'Water' will pick out whichever substance happens to play the watery role in their respective worlds. It just happens that *nothing* plays the role on Dry Earth. Doscar is, so to say, mislocating himself on the two-dimensional matrix: his practice of using 'water' tacitly assumes that his world is one of the countless possible ones where *some* substance plays the watery role. He is not, in his normal use, considering Dry Earth as actual. And because his term fails to pick out a substance on Dry Earth, it will similarly fail to have an extension in other possible worlds, considered as counterfactual. Nonetheless, Doscar's usage of 'water' will pick out H₂O on Earth, considered as *actual* by Doscar. And similarly for XYZ on Twin Earth.

This is, in my view, a sufficient ground for saying that Doscar's tokens of 'water' *do* express a determinate concept, although one which lacks an extension. To the extent that there are determinate answers to what substances *our* usage of 'water' would pick out in different possible worlds considered as actual, there are also answers to the same questions concerning *Doscar's* usage of 'water'. Accordingly, we can say that Doscar's tokens of 'Water is wet' will be false in all possible worlds considered as *counterfactual* by Doscar, but nonetheless their

content will be perfectly determinate, because semantic values can be assigned to the utterance in other possible worlds considered as *actual*.¹²

In his response to Moya, Boghossian parodies the answer by claiming that with the same strategy we could give truth conditions to the nonsense sentence 'mimsies are borogroves':

I take it that neither 'mimsy' nor 'borogrove' has meaning, so that someone uttering 'All mimsies are borogroves' doesn't say anything and his statement has no truth conditions. Mimicking Moya's suggestion, however, we seem to be able to say that it is meaningful after all. For consider:

'Mimsies are borogroves' is true iff mimsies are borogroves. Since neither 'mimsy' nor 'borogrove' refers to anything, we can take the view that 'Mimsies are borogroves' is false. What are its truth conditions? It would be true if there were some things in the extension of 'mimsy' and some things in the extension of 'borogrove' and if all the former were instances of the latter. And the concept we do express by 'mimsy' is the concept that we would express if there were in fact things in the extension of 'mimsy'. (Boghossian 1998, p. 260)

It is hard to take this response very seriously. Boghossian completely overlooks the fact that the use of 'water', on Dry Earth, is governed by conventions which make it the case that it *does* have an reference-fixing description associated with it, while 'mimsy' and 'borogrove' on Earth are not. That is why both our and Doscar's utterances of 'Water is wet' have semantic values in a range of possible worlds, considered as actual. Dry Earthlings have quite a clear idea of what would fall under the concept they express by 'water',¹³ whereas we have no idea at all of what a thing would have to be like in order to be properly called a 'mimsy' or a 'borogrove'.

The Dry Earthlings are, as noted earlier, in the same situation with their term 'water' as Earthlings were with 'phlogiston'. It *seems* to them that there is a local watery stuff on Dry Earth just as it seemed to Earthlings that there is a local phlogistony stuff. The concept DWATER, on Dry Earth, aims to denote this stuff, and because the Dry Earthlings assume that there is an underlying naturalistic essence to be found, the concept aims to denote a natural kind. But there is no such natural kind, not even a motley kind. Dry Earthlings are

consistently deceived by something or other in their environment. Yet, with the above strategy, we can show that Dry Earthlings do express a determinate concept with their tokens of 'water', and similarly for Earthlings and 'phlogiston' in the past.

VI

There are, as far as I can see, two kinds of worries one might raise about the appeal to two-dimensional semantics. These worries stem from two possible ways of understanding the role of primary intensions, giving rise to two different ways of understanding what it is to consider a possible world as actual (Stalnaker 2001). I will now briefly state the two worries and explain why neither of them is very serious.

On the first interpretation of primary intensions, the *semantic* interpretation (Stalnaker 2001, pp. 148-149), the primary intension (*i.e.* a rigidified description such as 'the actual watery stuff') is part of the meaning of the term 'water' or content of the concepts WATER, TWATER, and DWATER. On this view, a competent user of the term or concept will, implicitly or explicitly, know the primary intension of the term or concept. The worry here is that this appeal to primary intensions will commit one to *internalism* about content, thus undermining the whole argument as a demonstration of compatibilism. After all, if Oscar, Toscar and Doscar all associate the same primary intension with their uses of 'water', are we not forced to say that they have the *same* concept, after all?

I do not think this follows. On this interpretation, to say that the three Oscars associate the same primary intension with 'water' is just to say that they are alike in their internal aspects, as far as their (actual and counterfactual) usage of 'water' is concerned. And that is something we are assuming in any case, in constructing the Twin Earth and Dry Earth scenarios. In itself, it says nothing about how concepts and thoughts should be *individuated*. The two-dimensional framework does not by itself commit one to either internalism or externalism.

An externalist who opts for the semantic interpretation should say that concepts are *partly*, but not *wholly* individuated by the primary intensions associated with their application. They are also partly individuated by external factors – by which substance (etc.), if any, satisfies the reference-fixing description. Thus, Oscar has the concept WATER, referring to H₂O, Toscar the different concept TWATER, referring to XYZ, and Doscar yet another concept DWATER, referring to nothing. The fact that the primary intensions of these concepts are identical does not make the concepts themselves identical. Furthermore, the fact that primary intensions *are* associated with these concepts, including Doscar's DWATER, differentiates Doscar's term 'water' from Boghossian's 'mimsy' and 'borogrove' (lacking both reference *and* primary intensions), as well as the phlogiston theorists' 'phlogiston' (also referring to nothing, but having a primary intension).

On the second interpretation, the *metasemantic* interpretation (Stalnaker 2001, pp. 149-150), the 'primary intensions of expressions are derivative from the secondary intensions that those expressions have in the different possible worlds.' (p. 149) On this view, the primary intensions are not part of the meaning of terms or contents of concepts, but rather part of the story of how the terms and concepts get to have the secondary intensions, that is, extensions across possible worlds, that they do. On this interpretation, we do not require that a competent user of a term or a concept know the primary intensions, even implicitly. Some description of the form 'the colourless, odourless, thirst-quenching liquid flowing in the rivers ...' is part of our explanation, as semanticists, of how it is that Oscar's term 'water' gets to apply to H₂O in all possible worlds, Toscar's 'water' to XYZ, and Doscar's to nothing. The worry here is that, if Doscar is not associating the reference-fixing description with his 'water', we would not be able to assign a content to his DWATER concept, after all. That is, if it is only *we semanticists* who associate the description with his 'water', the only job the description would do is to

explain why his 'water' fails to refer, and it would not help in individuating the concept DWATER.

This second worry would only cause trouble if we were to hold that the referent (determined by the description) *exhausts* the content of WATER, TWATER, and DWATER. On such a view, the Dry Earth scenario would indeed be a problem: the intuitive difference between Doscar's 'water' and our 'mimsy', 'borogrove', or 'phlogiston' in the past would disappear. But there is no reason why an externalist opting for the metasemantic interpretation should have to hold this view.¹⁴ Even though we are not assuming competent users of terms and concepts to know the primary intensions in any sense, we must still assume that *something* about the subjects using those terms *does* make it the case that we, *as semanticists*, are justified in using those descriptions in explaining why the terms and concepts have the *secondary* intensions that they do. As long as this *something* enters, together with the referent, into the individuation of the meaning or concept in question, we can use the above strategy to show that Doscar's tokens of 'water' and 'water is wet' do have determinate contents, even though his 'water' fails to refer.¹⁵

Here's another way to make the point. Even if primary intensions are not known by speakers and thinkers, on the metasemantic interpretation something must make it the case that they *do* tell us what terms and concepts refer to. Suppose that we, as semanticists, know enough about subject S to know how the referent of his or her use of 'water' would depend on the context – that we know which primary intension to associate with the term – but we did not know enough about the context to actually work out the referent. Now, based on the way Twin Earth and Dry Earth are set up, this knowledge would not enable us to know whether S is Oscar, Toscar or Doscar. But, and this is the crucial point, even if S is Doscar, *there is something for the semanticist to know* about his use of 'water'. We can, in a sense, consider different worlds as actual, from a semanticist's detached viewpoint. Suppose S turns out to be

Doscar. Before we learned that S's environment is such that his 'water' fails to refer, we knew there *was* a way in which the referent *would* depend on the environment. But with 'mimsy' and 'borogrove' there is nothing of the sort for the semanticist to know, and this is enough to show that Doscar's 'water' is *not* like 'mimsy' or 'borogrove'.

VII

If the above is correct, an argument of the McKinsey type will not enable us to conclude, *a priori*, given externalism and self-knowledge, that water exists, or anything as implausible as that. As far as I can see, the strongest argument of this general form which *might* be supported by this kind of externalism would be something like the following:

Argument C:

- (1c) A(If I have concept K, and K aims to denote a natural kind, there is something which has appeared to satisfy the reference-fixing description associated with K.)
- (2c) A(I have concept K, and K aims to denote a natural kind.)
therefore,
- (3c) A(There is something which has appeared to satisfy the reference-fixing description associated with K.)

Premiss (1c) does not seem to be a direct consequence of externalism, but may be derivable from externalism together with suitable additional *a priori* premisses concerning experience and causation. I do not know whether such a derivation can be given, but even if we suppose it can, we should note that there is nothing very objectionable about (3c). We need to keep in mind that the 'something which has appeared to satisfy the reference-fixing description' need not be the referent of K. If, for example, K is PHLOGISTON, we can deduce, *a priori*, that something has appeared to be a substance released in combustion. But it does *not* follow that this something is *phlogiston*. It was part of the concept PHLOGISTON that the referent should be a *substance*, the local phlogistony *stuff*. That is (part of) why the concept *aimed* to denote a natural kind. When it turned out that combustion is not the process of releasing

anything, but rather that of combining with oxygen, the chemists of the time did not take this as the discovery that phlogiston is 'lack of oxygen' or anything of the sort. That would not have qualified as 'stuff' – hence, it could not be phlogiston. Rather, they took it as the discovery that *there is no phlogiston*.

Notice, moreover, that it is no longer crucial to an argument such as C that K be a natural kind concept. Any concept where our practices make it the case that a description fixes the extension of a concept will do. For example, it is customary to think that our actual colour judgements fix the extensions of our colour concepts. Whichever property it happens to be which in the actual world makes us see things as red – and surely this property is not a natural kind, but highly disjunctive – is also the property of redness in other possible worlds. Hence, the extension of the concept RED (our concept, that is) is partly fixed by external factors – by what property in fact happens to produce the appropriate sensations in us. It will then be the case (and knowable *a priori*) that if I have the concept RED, then there is something which has produced the sensations. Since I can also know *a priori* that I have the concept RED, I can deduce, *a priori*, that something exists which has produced the sensations. And similarly for many other concepts. But I maintain that there is nothing here that we should find very troubling. All we can deduce *a priori* from such arguments is that something is causing us to perceive various things – we do not get an *a priori* proof of anything more detailed than that.

Department of Philosophy
20014 University of Turku
Finland

jussi.haukioja@iki.fi

Bibliography

- Boghossian, P. (1989). 'Content and Self-Knowledge', *Philosophical Topics* 17, 5-26.
- Boghossian, P.A. (1997). 'What the Externalist Can Know a Priori', *Proceedings of the Aristotelian Society* 97, 161-175.
- Boghossian, P.A. (1998). 'Replies to Commentators', *Philosophical Issues 9: Concepts* (ed. E. Villanueva), 253-260.
- Brewer, B. (2000). 'Externalism and a Priori Knowledge of Empirical Facts', in P. Boghossian and C. Peacocke (eds.): *New Essays on the a Priori* (Oxford: Clarendon Press), 415-432.
- Brown, J. (1995). 'The Incompatibility of Anti-Individualism and Privileged Access', *Analysis* 55, 149-156.
- Brueckner, A. (1992). 'What an Anti-Individualist Knows a Priori', *Analysis* 52, 111-118.
- Brueckner, A. (2001). 'A Priori Knowledge of the World Is Not Easily Available', *Philosophical Studies* 104, 109-114.
- Chalmers, D. (2002). 'The Components of Content', in D. Chalmers (ed.): *Philosophy of Mind: Classical and Contemporary Readings* (Oxford: Oxford University Press), 608-633.
- Davies, M. (2000). 'Externalism and Armchair Knowledge', in P. Boghossian and C. Peacocke (eds.): *New Essays on the a Priori* (Oxford: Clarendon Press), 384-414.
- McKinsey, M. (1991). 'Anti-Individualism and Privileged Access', *Analysis* 51, 9-16.
- McLaughlin, B.P. and Tye, M. (1998a). 'Externalism, Twin Earth, and Self-Knowledge', in C. Wright, B.C. Smith and C. Macdonald (eds.): *Knowing Our Own Minds* (Oxford: Oxford University Press), 285-320.
- McLaughlin, B.P. and Tye, M. (1998b). 'Is Content-Externalism Compatible With Privileged Access?', *Philosophical Review* 107, 349-380.
- Miller, R.W. (1997). 'Externalist Self-Knowledge and the Scope of the a Priori', *Analysis* 57, 67-75.
- Moya, C.J. (1998). 'Boghossian's *Reductio* of Compatibilism', *Philosophical Issues 9: Concepts* (ed. E. Villanueva), 243-251.
- Sawyer, S. (1998). 'Privileged Access to the World', *Australasian Journal of Philosophy* 76, 523-533.
- Stalnaker, R. (2001). 'On Considering a Possible World as Actual', *Proceedings of the Aristotelian Society* Supplementary Volume 65, 141-156.
- Warfield, T.A. (1998). 'A Priori Knowledge of the World: Knowing the World by Knowing Our Minds', *Philosophical Studies* 92, 127-147.

1 An earlier version of this paper was presented at the Fourth European Congress of Analytic Philosophy, in Lund, Sweden (2002). I am grateful for the audience for discussion. This work has been financially supported by the Academy of Finland (projects 52379 and 202513).

2 *A priori* is here used in the weak sense of 'knowledge obtained independently of empirical investigation' (McKinsey 1991, p. 9).

3 Two other incompatibilist strategies are those based on so-called 'switching' -cases (cf. especially Boghossian 1989), where we assume that a subject is transported from Earth to Twin Earth or vice versa, and the disjunctive version of the McKinsey recipe (Brown 1995), where the proposition P (cf. below) is of the form 'instances of K exist or I am a part of a speech community where concept K is used'. These arguments have, in my opinion, been convincingly refuted by McLaughlin & Tye (1998a).

4 I use small capitals to denote concepts. 'Having a concept' is here to be taken just as the ability to form propositional states, in the content of which the concept figures. Whether we think of concepts as capacities or as mental particulars which ground our capacities is irrelevant here.

5 It should be noted that (A) does *not* say that it is *a priori* knowable that, if I have the concept of water, H_2O exists, but only that, if I have the concept of water, *water* exists. The extensional equivalence of water and H_2O is *a posteriori*, so the latter claim does not entail the former. Brueckner (1992) takes McKinsey to have meant the former claim and defends compatibilism by noting that no externalist would put forward such an absurd claim. However, the problem arises with the latter claim as well.

6 Attempts have also been made to answer the McKinsey recipe by denying the transition from (A)&(B) to (C). Davies (2000) attempts to establish that the reasoning $A(p \rightarrow q), A(p), \text{ therefore } A(q)$ is not always cogent ('A' stands for 'it is *a priori* that'). Miller (1997), on the other hand, claims that McKinsey equivocates on different senses of *a priori* knowledge (or knowability). I do not think either response works, but will not argue against them here. I believe that the McKinsey recipe can be refuted without resorting to these strategies.

7 Unacceptable to most people, but not quite all: see Sawyer (1998) and Brewer (2000).

8 The presumption is defeasible – we can, for example, allow that 'water' aims to denote a natural kind in worlds where the watery stuff in the extension of 'water' consists of a number of different substances.

9 Arguments of this form have been proposed as answers to (at least some forms of) skepticism; see, for example,

Warfield (1998). For an answer to Warfield based on the problem pointed out here, see Brueckner (2001).

10 Note that I am not assuming that all users of 'water' would associate the *same* reference-fixing description with the term – only that Oscar, Toscar and Doscar would, and this should be unobjectionable because of the way Twin Earth and Dry Earth are set up.

11 Of course, there are various versions of two-dimensional semantics. I am taking the terminology from Chalmers (2002): his treatment of the imaginary scenarios as *epistemic possibilities* is well suited to this context. However, I do not want to commit myself to his view that primary intensions are the *contents* of our concepts; see section VI.

12 If we do the two-dimensional semantics for 'Water exists' à la Chalmers, we will find that the sentence is not *a priori* in the two-dimensionalist sense of *a priori*. However, one should not make too much of this. The two-dimensionalist sense of apriority is not the sense of apriority relevant to the issue of compatibilism – I am not claiming that the contingency of the primary intension establishes compatibilism. Rather, I am invoking the two-dimensional machinery to show that Doscar's tokens of 'water' express a determinate concept. In fact, I share Stalnaker's (2001, 153) suspicion that two-dimensionalist 'apriority' may not have epistemological significance at all.

13 To be more precise: they know what would be in the extension of their 'water' in each of a vast array of ways the world could turn out for all they know *a priori*.

14 Cf. McLaughlin and Tye's (1998a, p.306n36) diagnosis that Boghossian may be falsely supposing that a Twin Earth externalist should hold a direct theory of reference.

15 To avoid side-tracking the discussion, I am being purposefully vague on what this *something* should be taken to be, whether it should be shared across speakers, and whether it is a strictly individual or partly social fact or feature.