

Externalism, Architecturalism, and Epistemic Warrant*

Martin Davies

This paper addresses a problem about epistemic warrant. The problem is posed by philosophical arguments for externalism about the contents of thoughts, and similarly by philosophical arguments for architecturalism about thinking, when these arguments are put together with a thesis of first person authority. In each case, first personal knowledge about our thoughts plus the kind of knowledge that is provided by a philosophical argument seem, together, to open an unacceptably ‘non-empirical’ route to knowledge of empirical facts. Furthermore, this unwelcome prospect of transferring a ‘non-empirical’ warrant from premises about our own mental states and about philosophical theory to a conclusion about external environment or internal architecture seems to depend upon little more than the possibility of knowledge by inference. (The use of the scare-quoted term ‘non-empirical’ is explained a couple of paragraphs further on.)

Introducing the Problem

For the purposes of this introduction to the problem, **externalism** about some mental property is the thesis that whether a person (or other physical being) has that property depends, not only on conditions inside the person’s skin, but also on the person’s environment and the way that the person is embedded in that environment. **Architecturalism** about some mental property is the thesis that whether a person (or other physical being) has that property depends, not only on the person’s (actual and counterfactual) patterns of behaviour, but also on the person’s internal cognitive architecture.

In each thesis, the dependence in question is supposed to be not causal but conceptual. It is the kind of dependence that can be revealed by philosophical theory, deploying arguments of a broadly a priori character. If the business of conceptual analysis is to develop the best elaborations and precisifications of the concepts that we actually use, then the methodology of externalist and architecturalist arguments will be conceptual analysis applied to the concept of having the mental property in question. (For a recent defence of conceptual analysis, and an account of its role in metaphysics, see Jackson, 1994, 1998.)

The problem about epistemic warrant arises when externalism or architecturalism is put together with a thesis of **first person authority**. For present purposes, a thesis of

* An early version of this material was given as an Invited Paper at the Eastern Division Meeting of the American Philosophical Association in December 1994; Joseph Owens commented and Kirk Ludwig chaired the session. Essentially the same version was presented at the conference held at St. Andrews in August 1995, on which this volume is based; Diana Raffman commented. The present version differs from that presented at the conference – the version on which Raffman’s comments (this volume) are based – in two main ways. There is a more thorough discussion of externalism in the lengthy Section 1, and there is some indication of the direction that future work might take in the brief Section 5.

I am grateful to the many colleagues and friends who have made comments and suggestions, or raised objections, including especially Bill Brewer, Jessica Brown, Graeme Forbes, Jay Garfield, Richard Holton, Frank Jackson, Wolfgang Künne, Rae Langton, David Lewis, Kirk Ludwig, Joseph Owens, Christopher Peacocke, Diana Raffman, Nigel Shardlow, Daniel Stoljar, Tom Stoneham and Mark Textor. Many of the points which they made will need to be addressed more adequately in future work.

Discussions at the Graduiertenkolleg Kognitionswissenschaft in Hamburg helped me to clarify the issues, and the paper was completed while I was a Visiting Fellow in the Research School of Social Sciences, Australian National University. I am grateful to both institutions.

first person authority about some mental property says that we have a distinctively first personal and specially authoritative way of knowing that we ourselves have that property, when we do have it, without needing to conduct any detailed empirical investigation either of the environment and our relation to it, or of our internal cognitive architecture. (We are not concerned here with differences that might be drawn between a thesis of first person authority and a thesis of privileged access. Indeed, the Privileged Access Thesis stated by Brian McLaughlin and Michael Tye (this volume, p. 286) could serve, with little modification, as a precisification of the idea of first person authority that is in play.)

We shall be considering a number of arguments of the following form (MC):

- (1) I have mental property M.
- (2) If I have mental property M then I meet condition C.

Therefore:

- (3) I meet condition C.

These arguments are potentially problematic to the extent that premise (1) is something that can be known with first person authority, while premise (2) is something that can be known by way of a philosophical argument for the general thesis:

(\Box x) (If x has mental property M then x meets condition C).

The scare-quoted term ‘non-empirical’ (used in the first paragraph) is intended to cover both the kind of epistemic warrant that attaches to first personal judgements about mental properties and the kind that attaches to the consequences of philosophical theory. The term is non-optimal in a number of ways, and perhaps ‘non-investigative’ would be less misleading; but let it stand. The rough idea is that both kinds of knowledge can be had without rising from the armchair. Given this use of the term, knowledge of premise (1) and of premise (2) is ‘non-empirical’; and it is obvious a priori that the conclusion (3) follows from those two premises. So, the transparently valid argument from (1) and (2) to (3) seems to offer a route to knowledge of (3), and with a warrant that is still ‘non-empirical’. But, in the particular arguments that we shall be considering, it is intuitively implausible that the conclusion (3) is something that can be known in a ‘non-empirical’ (non-investigative, armchair) way. Rather, knowledge of (3) would seem to require detailed empirical investigation of the world (whether outside the skin or inside the skull). This is the problem about epistemic warrant that is posed by externalist or by architecturalist arguments, when those arguments are put together with a thesis of first person authority. Is there any principled way of blocking the transfer of the ‘non-empirical’ warrant for (1) and (2) across the a priori known entailment to (3)?

This way of setting out the problem – at least as it concerns externalism about the contents of thoughts – owes a good deal to Michael McKinsey (1991). (McKinsey uses the term ‘a priori’ where I use ‘non-empirical’. McLaughlin and Tye (this volume) provide a detailed discussion of what they call ‘McKinsey’s recipe’.) McKinsey uses the problem to argue that externalism is incompatible with privileged access. If externalist philosophical theory delivers a ‘non-empirical’ warrant for premise (2), and the conclusion (3) cannot be known ‘non-empirically’, then privileged access does not extend to premise (1).

Critics of McKinsey’s argument often claim that, to the extent that the conclusion (3) is not something that could be known ‘non-empirically’, premise (2) cannot be known ‘non-empirically’ either, even given the correctness of externalism. (See Brueckner, 1992, for this kind of response.) Like these critics, I want to maintain that externalism is

compatible with first person authority (privileged access). But unlike these critics, I am not confident that it will always be possible to defuse a threatening McKinsey-style argument by claiming either that the conclusion (3) is so thin as to be knowable ‘non-empirically’ or else that knowledge of premise (2) depends on detailed empirical investigation of the world.

It is because I envisage a kind of worst-case scenario, in which this more usual line of compatibilist response is not available, that I focus on the question of whether there is a principled way to block the transfer of the ‘non-empirical’ warrant from premises to conclusion. Further motivation for this particular focus is then provided by the problem that is posed by architecturalist arguments (though it is, of course, perfectly possible to be an externalist about the content of thoughts without believing that there are any good architecturalist arguments that have a broadly a priori character).

Epistemic Warrant and Knowledge by Inference

Expressing the problem in terms of transfer of epistemic warrant might be misleading, for it might suggest that our project is to devise a cunning restriction on a general rule which otherwise permits properties of epistemic warrant to be transmitted across a priori known entailments.

In fact, there is no such general rule of transfer, for there are many properties of epistemic warrant which cannot be transferred across known entailments. The most obvious example is the property of being a kind of epistemic warrant that is non-inferential. But disjunctive properties of epistemic warrant would also raise questions about any general rule of transfer. Suppose that each of two properties, **F** and **G**, of epistemic warrants can be transferred across known entailments, in the sense that if all the premises have warrants with the property then the conclusion also has a warrant with that property. And suppose that some of the premises of an argument have epistemic warrants with property **F**, while the other premises have epistemic warrants with property **G**. Then each of the premises has a warrant with the disjunctive property **F or G**; but there is no reason to suppose that the epistemic warrant for the conclusion has either property **F** or property **G** (since not all the premises have warrants with property **F** and not all the premises have warrants with property **G**). So there is no reason to suppose that the disjunctive property, **F or G**, of warrants can be transferred.

Consider an example. Perhaps, if all the premises of an inference are known a priori then the conclusion can be known a priori. And perhaps, if all the premises of a sufficiently obvious inference are known with first person authority then the conclusion can be known with first person authority. But, it would not follow from those suppositions about transfer of warrant that if one of two premises is known with first person authority, and so is known either a priori or with first person authority, and the other premise is known a priori, and so is also known either a priori or with first person authority, then the conclusion is known either a priori or with first person authority.

This example might, for a moment, encourage the thought that we already have a solution to our epistemological problem in the fact that ‘non-empirical’ warrants are supposed to include both the a priori and the first personal. But, in fact, considering the prospects for general rules of transfer of warrant is not the best way to see what the nature of the problem is. It is more helpful to consider that ‘non-empirical’ is a negative notion. For the arguments that we shall consider, I do not need to engage in any detailed empirical investigation of the world (outside or within) in order to know premise (1); I just need to introspect (in a suitably neutral use of that term). Nor do I need to engage in any detailed empirical investigation in order to know premise (2), I just need to follow

through a philosophical argument. I do not need to conduct any empirical investigation to draw the conclusion (3) from the premises (1) and (2); the validity of the instance of modus ponens is obvious. So, unless there is some restriction on the possibility of knowledge by inference, by drawing that conclusion I gain knowledge of (3) – and still without any detailed empirical investigation.

The problem posed by the arguments that we shall be considering is that they seem to call for some restriction on knowledge by inference.

Knowledge by Inference and Deductive Closure

We can distinguish between questions about knowledge by inference and questions about the closure of knowledge under known entailment. A question about closure would be whether I can know premise (1) and know premise (2) but fail to know, or to be in a position to know, the conclusion (3). Our question about knowledge by inference is whether I can know premise (1) and know premise (2) but fail to be in a position *thereby* to know the conclusion (3).

Expressing the two questions in this way makes it clear that, in principle, there could be some restriction on knowledge by inference without there being any failure of closure. Knowing the premises (1) and (2), one might be in a position to know the conclusion (3), but not to know it by way of knowing those premises. It might even be that anyone who knew those premises would also know the conclusion, even though knowledge of the conclusion could not be arrived at by inference from those premises.

Of course, if knowledge is not closed under known entailment – in the sense that someone could know certain premises without being in a position to know a conclusion which she knows to be entailed by those premises – then a fortiori someone could know those premises without thereby being in a position to know the conclusion; so there must be some restriction on the scope of knowledge by inference.

It appears, then, that while any failure of closure requires some restriction on knowledge by inference, it might be possible to restrict knowledge by inference without breaching closure. But it would be reasonable to suppose that, in the context of the arguments that we shall be considering, the strategy that we just mentioned for securing a restriction on knowledge by inference without denying closure may not be a plausible one. So we should assume (pending detailed discussion) that restricting knowledge by inference will involve denying closure. I shall briefly mention, at the end of the paper (Section 5.2), how the epistemic situation can be redescribed in such a way as to honour the closure principle, by invoking the idea of context shifting (DeRose, 1995; Lewis, 1996).

Outline

In Section 1, I explain in some detail how the epistemological problem is posed by arguments for externalism about the contents of thoughts. In Section 2, I turn, much more briefly, to the apparently parallel problem posed by architecturalist arguments. In Section 3, we begin our journey towards a solution by considering some arguments that are superficially of the right form to be anti-sceptical arguments, but which cannot be used for anti-sceptical purposes. This leads to a principle for limiting the transfer of epistemic warrant across even a priori known entailments. However, this first limitation principle does not itself provide a solution to the problems that arise from externalist and architecturalist arguments. In Section 4, I introduce a second limitation principle on transfer of epistemic warrant – a principle that does close off the unwelcome prospect of

a ‘non-empirical’ route to knowledge of empirical facts. The final section indicates some possible directions for future work.

1. *Externalism*

We have said that externalism about a mental property, *M*, is the thesis that whether or not a person (or other physical being) has *M* depends, not only on conditions inside the person’s skin, but also on the person’s environment and the way in which the person is embedded in that environment. It may be helpful, now, to distinguish between two kinds of externalist thesis: *constitutive externalism* and *modal externalism*.

1.1 *Constitutive and Modal Theses*

Constitutive externalism and its opposite, constitutive individualism, are theses about the kind of philosophical account that can be given of what it is for an individual to have a certain property. Constitutive individualism (as it concerns mental property *M*) says that the fundamental philosophical account of what it is for an individual to have *M* does not need to advert to the individual’s environment (whether physical or social), but only to what goes on within the spatial and temporal boundaries of the individual. Constitutive externalism says that the fundamental philosophical account of what it is for an individual to have *M* does need to advert to the individual’s physical or social environment. (We shall not be concerned with social externalist theses (Burge, 1979) here.)

Modal externalism and its opposite, modal individualism, are theses about the existence or otherwise of Twin Earth examples. A Twin Earth example for a property exhibits a person (or other physical being) in an environment, and a duplicate person (or other physical being) in a second environment, where the first person has the property in question, and the second person does not, despite being the same as the first person from the skin inwards. A modal externalist thesis (concerning mental property *M*) says that there are Twin Earth examples for *M*. Modal externalist theses differ in modal strength along various dimensions, according as the two environments are allowed to be two parts of the actual world, or two parts of some single possible world, or parts of different possible worlds one of which is the actual world, or parts of two possible worlds neither of which need be actual. A modal individualist thesis, being the opposite of a modal externalist thesis, denies the existence of Twin Earth examples for *M*. Modal individualist theses are local supervenience claims varying, once again, in modal strength.

It would not be quite correct to say that constitutive individualism, as defined, entails modal individualism. The reason is that an individual may have properties (impure intrinsic properties) whose specification does not advert to anything outside the skin of that individual, yet which are not shared by duplicates. Cases in point, for an individual *a*, would include the property of being identical with *a*, and the property of having *a*’s brain as a proper part. Even if *a* and *b* are duplicates, *a* has, but *b* lacks, the property of being identical with *a*, and the property of having *a*’s brain as a proper part. So, it is a straightforward matter to provide Twin Earth examples for these properties – and without even needing to have the second environment different from the first. These cases do not involve mental properties; but it is clear that there are mental properties which can, *prima facie*, be specified without talking about anything outside the individual’s skin, but which would not be preserved across duplicates.¹

¹ This paragraph corrects a claim made in earlier work about the relationship between constitutive and modal individualism (Davies, 1991b, 1992b, 1993, 1996). I am grateful to Richard Holton and Rae Langton for discussion of this issue. For the distinction between pure and impure properties, see Langton, 1993; Langton and Lewis, 1998; Humberstone, 1996; Khamara, 1988.

It is a good question whether what is really at issue in debates about externalism and individualism is best characterised in terms of properties that can be specified without adverting to the environment or in terms of properties that are preserved across duplicates. But for present purposes, we shall simplify the discussion by treating constitutive individualism (concerning mental property *M*) as saying that the fundamental philosophical account of what it is for an individual to have *M* does not need to advert to the individual's physical or social environment, but only to properties of the individual that are shared by duplicates. (That is, we simplify by ignoring impure intrinsic properties.) With this simplification, it becomes trivial that constitutive individualism entails modal individualism, of whatever modal strength: supervenience holds and Twin Earth examples do not exist.

Likewise, modal externalist theses (of whatever modal strength) entail constitutive externalism. If there is a Twin Earth example for mental property *M*, and if that example works by placing the duplicates in different environments (and does not trade on differences in impure intrinsic properties), then the fundamental philosophical account of what it is for an individual to have *M* must advert to the individual's environment. However, the converse entailment is not guaranteed to hold: a property that is canonically specified in externalist terms might yet be preserved across duplicates. Frank Jackson and Philip Pettit (1993, p. 271) give the example of dispositional properties. An account of the property of being soluble in water will surely advert to conditions external to any individual thing (such as a sugar cube) that has that property. But still, if this sugar cube, *c*, is water soluble then any duplicate sugar cube, *d*, is also water soluble, even if *d* is located in an environment devoid of water. (Perhaps we need to restrict attention to nomologically possible environments; but this complication does not matter for present illustrative purposes.)

Someone might conceive of cognitive abilities, like the ability to have thoughts of one or another type, as analogous to solubility. Having the ability to think about water (to think water thoughts) would be conceived of as a constitutively, but not modally, externalist property of thinkers. Having the ability to think about perceptually presented sugar cubes might be conceived of in the same way – as a property shared by duplicates. On this conception, it would be allowed that one thinker, *a*, might deploy this ability when presented with sugar cube *c*, and so think about *c*, while a duplicate thinker, *b*, thought instead about sugar cube *d*. So, the mental property of thinking about sugar cube *c*, in particular, would be both constitutively and modally externalist; while the mental property of being able to think about sugar cubes would be constitutively, but not modally, externalist.

Someone favouring this conception of cognitive abilities might take a further step. Perhaps *a* thinks, 'That sugar cube is old', thereby thinking a thought that is true just in case *c* is old, while *b* thinks, 'That sugar cube is old', thereby thinking a thought that is true just in case *d* is old. Perhaps *a*'s thought is true while *b*'s thought is false: they think different particular thoughts. But still, it might be said, they think thoughts of the same type (the 'that sugar cube is old' type) and the mental property of thinking a thought of that type is shared by duplicates, even though it is a constitutively externalist property.

Once that step has been taken, it may be extremely tempting to describe the situation in the following way. Thinking, concerning sugar cube *c*, that that sugar cube is old is really a composite matter. There is a genuinely mental component, which is the deployment of a cognitive ability; and there is an environmental relational component, which is the presence, near the thinker, of sugar cube *c*. The thinkers, *a* and *b*, are just the same in respect of the genuinely mental aspect of their thinking; they differ only in that *a*

is environmentally related to **c** in the way that **b** is related to **d**. So, we might just as well say that **a** and **b** think thoughts with the same *content*, even though they think *about* different sugar cubes. So far as the content (properly so-called) of the thoughts goes, modal externalism has not been shown to be correct. A Twin Earth example might demonstrate modal externalism for the aboutness of the thoughts; but this is not of very much interest if aboutness is a matter of environmental relational properties that are wholly extrinsic to the content of the thoughts. So it might be said. (This use of the term ‘aboutness’ to contrast with ‘content’ should be distinguished from the use in which aboutness is more or less the same as intentionality.)

If the situation is described in this way, then it becomes difficult to see how the form of argument (MC):

- (1) I have mental property M.
- (2) If I have mental property M then I meet condition C.

Therefore:

- (3) I meet condition C.

could lead to the kind of epistemic problem that we have described. Premise (1) is supposed to be known with first person authority; but first person authority plausibly extends only to the content, and not to the (environmental relational) aboutness, of thoughts. On the other hand, premise (2) is supposed to follow from a general (and not merely contingent) thesis:

- $$(\Box x) (\text{If } x \text{ has mental property M then } x \text{ meets condition C})$$

that would support modal externalism. (If an individual *x* does not meet (environmental) condition C, then *x* does not have mental property M, even if *x* is a duplicate of me.) But modal externalism seems (so far) to apply to the aboutness properties, and not to the content properties, of thoughts.

We need to have some idea of a range of externalist examples in which, on the one hand, the externalism concerns the content of thoughts – content that is within the range of first person authority – and yet, on the other hand, the externalism is modal as well as constitutive. But before we turn to some of the varieties of externalist examples (Section 1.3), we must deal with one further question. How is the externalist supposed to argue for premise (2)?

1.2 Arguing for Premise (2)

This question is potentially puzzling when asked against the background of the distinction between constitutive and modal externalist theses. Although there is no guarantee that every constitutive externalist thesis entails modal externalism, still, it might be possible to argue for premise (2) via a constitutive modal thesis in a specific case. But the problem here is that a thesis about the fundamental philosophical account of the mental property of having thoughts with certain contents would seem to depend on a detailed theory of intentionality. So, in the absence of any agreed theory of intentionality it is little wonder that actual externalist arguments typically make use of Twin Earth examples; even if they are aimed, ultimately, at a constitutive externalist conclusion, they go via modal externalism.

But, if premise (2) is the target, then arguments from Twin Earth examples have a disadvantage in that they do not make it clear exactly what the commitments of externalism are. They are liable to underestimate the environmental requirements for having a thought with a certain content. An individual *x* has mental property M in

possible world w_1 . The environmental differences between w_1 and an alternative possible world, w_2 , are supposed to be enough to secure that y in w_2 , though a duplicate of x in w_1 , differs from x in respect of property M . Typically, the aim is not merely to show that y lacks M in w_2 , but also to exhibit y as having some determinate different mental property M' in w_2 . Now, in order that the Twin Earth example should carry conviction even while there is no agreed theory of intentionality, it is likely to involve inter-world differences that go well beyond what is strictly speaking required according to whatever theory of intentionality turns out to be correct. In particular, the description of possible world w_2 may well pile on respect after respect in which w_2 differs from w_1 , in order to provide a plausibly sufficient condition for y to have M' rather than M . But then, when we come to read off something of the form:

($\Box x$) (If x has mental property M then x meets condition C)

from the example, all that we can fill in for ‘condition C ’ – a necessary condition for having M – is the negation of a perhaps quite lengthy conjunction of features of y ’s environment in w_2 .

We can make some progress with this problem, and offer the prospect of a more substantial necessary condition in premise (2), if we note that there is no need for a Twin Earth example to do any more than make it plausible that y lacks M in w_2 . To the extent that the sufficient condition for lacking M is thinner than that for having M' , it will yield, under negation, a thicker substitution for ‘condition C ’. Or, if we consider again arguments that proceed via constitutive externalist theses, we can note that, in order to argue for premise (2), we do not, in fact, require a complete theory of intentionality. We only need to spell out necessary conditions, and not sufficient conditions, for a thinker to have a thoughts with certain contents.

Even with these points noted, however, someone might urge that, to the extent that philosophical theory uncovers externalist necessary conditions for having thoughts with certain contents, these conditions will be so thin that there will be nothing counterintuitive about the prospect of being able to know, from the armchair, that we meet them. To this I offer two partial replies. The first is (to repeat a point made earlier) that, since I want to defend both externalism and first person authority, and since we are as yet some distance from agreeing whether certain conditions are necessary for a thinker to have thoughts with this or that content, I would prefer to be prepared for a worst-case scenario. The second partial reply consists in sketching some externalist examples.

1.3 Varieties of Externalist Examples

Let us return to the contrast between content and aboutness and the need for some modal externalist examples that decisively concern content. At the beginning of ‘Other Bodies’ (1982), Tyler Burge comments on a case that is similar to our example of the thinkers **a** and **b** and the sugar cubes, but involves a single thinker, Alfred, thinking about two apples (1982, p. 97):

We may say that Alfred has the *same* belief-content in both situations. It is just that he would be making contextually different applications of that content to different entities. His belief is true of apple 1 and false of apple 2. The *nature* of his mental state is the same. He simply bears different relations to his environment. . . .

This deflationary interpretation seems to me to be correct.

So Burge, like our imagined theorist who started out from the analogy with solubility, does not regard Twin Earth examples that work by replacing an object that is thought about (apple 1 or sugar cube **c**) with a different object (apple 2 or sugar cube **d**) as demonstrating externalism about content.

In Burge's writings, there are at least three other kinds of Twin Earth example. There are examples that are intended to illustrate social externalism, such as the 'arthritis' and 'sofa' examples (Burge, 1979). There are examples that are intended to illustrate externalism about the content of perceptual experiences, such as the example of the shadows and the cracks (Burge, 1986). And there are examples that are intended to illustrate externalism about natural kind concepts (particularly as those concepts are exercised by thinkers who lack relevant scientific knowledge) such as the 'water'/'twater' example (Burge, 1982; Putnam, 1975). The first kind of example is not strictly relevant to our present purposes, since we are setting aside the question of social externalism. The second kind of example is potentially problematic because it concerns the non-conceptual content of experiences. Although it is possible to provide convincing Twin Earth examples of this type (in my view: Davies, 1991b, 1992b, 1996), the examples involve imagined systems or creatures of great simplicity (Davies, 1993) and this simplicity counts against the idea of concept deployment. The third kind of Twin Earth example, involving natural kind concepts, is certainly relevant. But, despite the fact that this is often taken to be externalism's most favoured case, it introduces a number of difficult and complex issues.

Externalism and natural kind concepts

Davies and Humberstone (1980) pointed out, in effect, that much of what Putnam (1975) said about the natural kind term 'water' would be consistent with the term's having the same sense as a definite description: 'the chemical kind to which that liquid belongs which actually falls from clouds, flows in rivers, is drinkable, colourless, odourless . . .' (1980, p. 18). But if thoughts that are expressed and reported by using the word 'water' are really thoughts in which some such definite description concept is deployed, then the contrast between content and aboutness can be invoked once again. The fact that the chemical kind H₂O fits the description is arguably a matter that is quite extrinsic to the content of the thought. (This conclusion does not depend on our ignoring the presence of the 'actually' in the description. See also Evans, 1979; 1982, p. 50.)

It is vital, then, that natural kind concepts, if they are to figure in externalist examples that genuinely concern content, should not be definite description concepts. Let us suppose that this is so, and move on to ask what might be the environmental relational necessary conditions for a thinker to have thoughts in which, say, the natural kind concept *water* is deployed. Burge says (1982, p. 114):

We want to say that it is logically possible for an individual to have beliefs involving the concept of water . . . even though there is no water . . . of which the individual holds these beliefs. . . .

I think we also want to say something stronger: it is logically possible for an individual to have beliefs involving the concept of water . . . , even though there exists no water.

So, on Burge's account, no encounters between a thinker and any samples of water are required if the thinker is to think, say, that water is wet; indeed, no water need exist in the thinker's world at all. The concept would remain available to a thinker – even a thinker who showed 'relative ignorance and indifference about the nature of water' – because we (other members of the thinker's community) 'would still have our chemical analyses' (Burge, 1982, p. 116).

The only entailment to which Burge is explicitly committed is this (1982, p. 117): 'In some instances, an individual's having certain *de dicto* attitudes entails the existence of entities other than himself and his attitude contents.' That raises the question whether externalism can generate good anti-sceptical arguments of the **(MC)** form – a question that will not be addressed directly here (though it is implicit in what is said here that it cannot). But we can ask whether any more specific entailments might be forthcoming.

On Burge's own account, where social externalism is in play, it might well seem that:

(\Box x) (If x has thoughts involving the concept of water then either there is water in x's environment or else there are other members of x's community).

If social externalism is set aside, then we might even have:

(\Box x) (If x has thoughts involving the concept of water then there is water in x's environment).

Jessica Brown considers the following 'entailment between mind and world' (1995a, p. 152):

Q: Necessarily, if x has a thought involving the concept of a natural kind K and x is agnostic about the application conditions of the concept of K, then either x is in an environment which contains K, or x is part of a community with the concept of K.

She argues that this is something to which Burge is committed. So, if Burge's philosophical theory is correct, then Q is something that could be known 'non-empirically'. Furthermore, it is plausible to suppose that a thinker can know 'non-empirically' that he is agnostic about the application conditions of a natural kind concept: he might know, for example, that he knows nothing about the chemical analysis of water. But Brown suggests that this entailment still cannot readily be used in an argument of the **(MC)** form so as to create an epistemological problem, since – at least by Burge's lights – 'a subject could be mistaken about whether a concept is a natural kind concept and still have that concept' (1995a, p. 154).

We might try taking the matter further by considering, alongside Q, a corresponding entailment concerning non-natural kind concepts (Brown, 1995a, p. 154):

R: Necessarily, if x has a thought involving a non-natural kind concept, C, and x is agnostic about the application conditions of C, then x is part of a community that has the concept C.

with a view to constructing an argument of the **(MC)** form with conclusion:

Either I am in an environment that contains instances of water and *water* is a natural kind concept or I am part of a community which has the concept *water* (whether or not *water* is a natural kind concept).

But Brown herself (1995b) raises a difficulty for this strategy; namely, that the application conditions of a non-natural kind concept may leave it indeterminate whether the concept applies in certain cases. Someone could have complete understanding of such

a concept and still be unsure about whether the concept applied to one or another object. As a result, it is not easy to maintain that a thinker can know, just by reflection, that he is agnostic about the application conditions of a non-natural kind concept (Brown, 1995b, pp. 215-6; see also McLaughlin and Tye, this volume, p. 317).

The overall situation concerning externalism and natural kind concepts is unclear. Suppose that we grant, for a moment, the claim that a subject could be mistaken about whether a concept is a natural kind concept and still have that concept. Then, one point to note is that this seems to be consistent with the idea that a subject who is not mistaken in judging that the concept *water* is a natural kind concept has knowledge of that fact – and perhaps even ‘non-empirical’ knowledge. But, second, we might question whether a subject could be mistaken about whether a concept is a natural kind concept. If social externalism is set aside, then this claim might look less plausible. Many issues would need to be settled before we could be confident about what is the right thing to say. But there seems to be at least the prospect of a ‘non-empirically’ knowable entailment along the lines of:

(\Box x) (If x has thoughts involving the concept of water then there is water in x’s environment);

or even:

(\Box x) (If x has thoughts involving the concept of water then x has had such-and-such encounters with water).

And such an entailment could be used to construct an argument of the **(MC)** form. So we had better be prepared.

Externalism and object-dependent thoughts

There are two rather different kinds of case in which someone might draw a contrast between content and aboutness. Consider first a thought that involves a definite description concept, a thought of the form ‘The F is H’. Perhaps **c** is the F; so the thought is about **c**. Perhaps **d** might have been the F; in that case the thought would have been about **d**. But, it might be said, the content of the thought would be the same in either case.

Consider second a thought that involves a perceptual demonstrative concept (or mode of presentation), a thought of the form ‘That G is H’. If **c** is the perceptually presented G, then the thought is about **c**. If **d** had been the perceptually presented G then the thought would have been about **d**. But, it might be said, the content of the thought would be the same in either case.

There is a conception of the content of a thought, familiar from the work of Gareth Evans (1982) and John McDowell (e.g. 1984, 1986), according to which the first kind of case does, but the second kind of case does not, allow a contrast between content and aboutness to be drawn in the way described. On this conception, thought is essentially representational, and the content of a thought specifies how the world would have to be for the thought to be correct. Furthermore, the thinker, in having the thought, knows how the world would have to be for the thought to be correct. As McDowell puts it, (1986, p. 140): ‘The underlying idea is that to entertain a proposition one must know how one’s thinking represents things as being.’

In the first kind of case, if we abstract away from the question of which object is the F, there is still a condition for the correctness of the thought; namely, that there should be a unique object which is F and that whichever object is (uniquely) F should be H. Furthermore, this condition for correctness is a condition that the thinker grasps just by

having the thought. In the second kind of case, however, the situation is very different. If we abstract away from the question of which object is being perceived and thereby being thought about, then we abstract away from the question of which object needs to be H in order for the thought to be correct. There is, of course, a condition for correctness that we, as theorists, can frame. There should be a unique object that is G and that is being perceived (and perhaps attended to) by the thinker, and this object should be H. But this is not a condition that the thinker grasps just by having the thought. For the thinker is not thinking about himself, or about perception (or attention); the thinker is just thinking that that G is H.

On this conception, we cannot separate content from aboutness in the second kind of case. The content of the perceptual demonstrative thought is dependent on the object that the thought is about. If the object were different, then the content would be different; if there were no object, then there would be no content (no way that the world is represented in thought as being). It is part of this conception that, in order to have a thought that is object-dependent in this way, the thinker must, in some sense, know which object is in question. (This is what Evans calls ‘Russell’s Principle’; 1982, p. 65.) This does not mean, of course, that in order to think, concerning a perceptually presented man, ‘That man is tired’, I need to know whether the man before me is Frank or Ernest. Indeed, just what is required, in order to meet the ‘know which’ requirement, is not easy to specify. As Evans himself says (1982, p. 89), ‘The difficulty with Russell’s Principle has always been to explain what it means’. But however exactly it is cashed out, the ‘know which’ requirement does – it seems to me – have a consequence that is of considerable importance for the purposes of this paper. Given that requirement, and the more general idea that ‘one must know how one’s thinking represents things as being’, it would be unmotivated to insist that, in the case of object-dependent thoughts, first person authority extends only to some (not fully representational) aspect of content that abstracts away from the object that is being thought about.

Someone who favours this conception of content has to disagree with Burge over the ‘deflationary interpretation’ of putative Twin Earth examples that involve perceptual demonstrative thoughts. On this conception, my thought, ‘That man is tired’, can be the basis for an argument of the **(MC)** form (cf. McLaughlin and Tye, this volume, p. 292).

Externalism and indexical thoughts

Much of what we have said about perceptual demonstrative thoughts goes also for indexical thoughts. When someone thinks, ‘The treasure is buried here’, or thinks while lying in bed in the dark, ‘There’s a bottle of whisky there’ (Evans’s example; 1982, p. 161), the content of the thought is place-dependent. A duplicate subject located at a different place would think a thought with a different content. To acknowledge this is not to blind ourselves to the similarity between the thoughts entertained by the two subjects. The two thoughts, with their different contents, are intuitively of the same type; and it is a requirement on any theory of these matters that it should give a proper account of that common type.

As before, it seems plausible to say that first person authority extends to the content of such an indexical thought. Given Russell’s Principle, it would be unmotivated to claim that the thinker’s first personal knowledge extends only to the type of the thought, and not to which place is in question. And we can easily construct a Twin Earth example, by having a thinker x located at a place p_1 and a duplicate thinker y located at place p_2 . But it is not clear that we can use these resources to develop an epistemologically problematic argument of the **(MC)** form. Suppose that I am located at p_1 . I may think that the

treasure is buried here and may know with first person authority that I think that the treasure is buried here. Furthermore, I may know as a piece of philosophical theory that in order to think a ‘here’ thought about a place a thinker needs to be located at that place. But I cannot know ‘non-empirically’ that in order to think that the treasure is buried here, I need to be located at p_1 and not at p_2 , say. For I cannot know ‘non-empirically’ that p_1 is here. So, the only conclusion that looks to be available from an argument of the **(MC)** form is that I am located here; and there is nothing problematic about the idea that I might know that ‘non-empirically’.

However, we need to consider the fact that being able to think about a particular place is not a trivial matter (Evans, 1982, p. 161):

We are prepared to suppose that there is a determinate thought here – that the subject has a definite place in mind – because we know that subjects do have a capacity to select one position in egocentric space, and to maintain a stable dispositional connection with it. . . . If the subject . . . does know which place his thought concerns . . . this will be manifestable only in manifestations of that stable dispositional connection . . .

What this suggests is that someone who is unable, for a while, either to maintain a stable dispositional connection with a position or to keep track of his movement through space is likewise unable, for that while, to have (determinate) indexical thoughts about places.

Evans himself presents an example that we can adapt (1982, p. 201):

A person might lie in bed in hospital thinking repeatedly ‘How hot it was here yesterday’ – supposing himself to be stationary in the dark. But his bed might be very well oiled, and be pulled by strings, so that every time he has what he takes to be the same thought, he is in fact thinking of a different place, and having a different thought.

As Evans describes the case, this thinker has several instantaneous thoughts about different places. But we can imagine that the person thinks, slowly, carefully, not wanting to knock anything over in the dark, ‘There’s a bottle of whisky just here’. Suppose that, during the time that it takes him to compose this putative thought, he moves several yards. Then it seems that we could reasonably deny that the subject manages any determinate thought at all – deny that the subject has a definite place in mind.

It might have been expected that, while the mental property of thinking a ‘here’ thought about a particular place, p_1 , would be both constitutively and modally externalist, only a thesis of constitutive externalism would hold for the property of thinking a thought of the ‘here’ type. But these last reflections suggest that we might be able to construct a Twin Earth example for that property. In one possible world there is x , stationary in bed in the dark, thinking, ‘There’s a bottle of whisky just here’ – a thought that is correct if there is indeed a bottle of whisky located at a particular position just next to the bed. In the alternative possible world there is y , moving silently along darkened corridors, and failing to think any ‘here’ thought at all.

We seem to have the resources for an epistemologically problematic argument of the **(MC)** form; and they are resources that are available even to someone who adopts a ‘deflationary interpretation’ of Twin Earth examples that rely on object-dependent or place-dependent thoughts.

We have been considering three kinds of externalist example that concern the contents of thoughts: examples of thoughts involving natural kind concepts, of object-dependent thoughts, and of indexical thoughts and thought types. None of these cases is

uncontroversial. But perhaps the discussion in this subsection has been sufficient to suggest that the prospect of an epistemological problem posed by arguments of the **(MC)** form is not wholly illusory.

1.4 Externalist Entailments and the Problem of Epistemic Warrant

The arguments of the **(MC)** form that concern us here make use (at premise (2)) of entailments issuing from externalist philosophical theory. These externalist entailments, in their generalised versions, are of the form:

ExtEnt($\Box x$) (If x believes that p then $E(x)$)

where ‘ $E(x)$ ’ is some statement about x ’s environmental embedding. An example of the relevant form, involving a natural kind concept, would be:

($\Box x$) (If x believes that water is wet then x has had such-and-such encounters with water)

Other examples, involving object-dependent thoughts or indexical thoughts, could be provided.

We assume that the methodology of philosophical theory is broadly a priori in character, and that the aim of philosophical theory is conceptual analysis, that is, the development of the best elaborations and precisifications of the concepts that we actually use. Thus, externalist philosophical theory aims, inter alia, at elaboration and precisification of the concept of believing, for example, that water is wet – the concept that we actually use in our judgements to the effect that we ourselves, or that others, believe that water is wet.

The potentially problematic arguments are of the form **Ext(MC)**:

Ext(1) I believe that p .

Ext(2) If I believe that p then $E(\text{me})$.

Therefore:

Ext(3) $E(\text{me})$.

I do not need to engage in any detailed empirical investigation of my environment to know premise Ext(1); for example, to know that I believe that water is wet. Nor do I need to conduct any such investigation in order to know the conditional premise Ext(2); for example, to know that if I believe that water is wet then I have had such-and-such encounters with water. And the conclusion Ext(3) obviously follows from these two premises. Yet it is implausible that I can know, without empirical investigation (from the armchair), the conclusion Ext(3); for example, can know ‘non-empirically’ that I have had such-and-such encounters with water. This is our epistemological problem, as it is posed by philosophical arguments for externalism about the contents of thoughts when these are put together with a thesis of first person authority about the contents of thoughts.

Although we assume that philosophical theory is broadly a priori, there is no need for us to be purists about this. There may be some empirical, a posteriori, assumptions in externalist theories. Suppose we consider the conditional made up of the conjunction of these assumptions as antecedent, and the conclusion Ext(3) as consequent. Then the epistemological problem remains if it is implausible that this conditional can be known in a ‘non-empirical’ way.

If externalism is to allow for first person authority, then we need some principled limitation on the transfer of epistemic warrant from Ext(1) and Ext(2) to Ext(3). We need to specify some restriction on knowledge by inference.

We might call our problem the *consequence problem* for first person authority given externalism. It should be distinguished from a different problem, which we can call the *achievement problem* for first person authority given externalism. How can I achieve a specially authoritative kind of knowledge about my own mental states, given that my being in those mental states depends on my environmental relational properties? For I am not specially authoritative about those properties.

1.5 *The Achievement Problem for First Person Authority*

Approaches to this problem typically make use of the fact that the content of my second-order belief – my belief that I believe that water is wet, for example – is dependent on the environment in just the same way as the content of my first-order belief – my belief that water is wet.

Knowledge is an achievement, and so there will be conditions that must be met if my second-order belief is to amount to knowledge. But there is no special problem in the fact that the content of the first-order belief is partly externalistically determined. My second-order belief, like my first-order belief, involves the natural kind concept *water*. So, in order for me to have that second-order belief at all, I need to have certain environmental relational properties; and if I do stand in those relations to my environment, then I meet the externalist conditions for having the first-order thought.

Consider, for example, a reliabilist account of knowledge, and suppose, for a moment, that we think in a language of thought. Then we can imagine the transition from the first-order belief to the second-order belief to be subserved by a causal mechanism that takes a language of thought sentence, *s*, as input, and produces a sentence, *O*^s, as output, where *O* is a language of thought sentential operator meaning ‘I believe that ___’. Because the externalist conditions that must be met if a language of thought sentence, *s*, is to have the content that water is wet are the same as those that must be met if *O*^s is to have the content that I believe that water is wet, this mechanism reliably produces true second-order beliefs. So, given reliabilism, those second-order beliefs count as knowledge. (Somewhat similarly, externalism is no impediment to the self-verifying character of the thought ‘I think, with this very thought, that water is wet’. See Burge, 1988.)

Alvin Goldman (1993) considers this mechanical kind of account of first personal knowledge of mental states, and complains that (p. 22) ‘it would leave the process of mental-state classification a complete mystery’. The complaint here is not an epistemological one. Rather, the worry is that the subpersonal level story about a mechanism operating on sentences of the language of thought does not tell us how to conceive of these transitions at the personal level. To deal with this worry, Goldman adds to the account ‘a plausible-looking hypothesis’, namely, that ‘mental states are states having a phenomenology’ (1993, p. 24).

This is a potentially problematic move to the extent that it suggests a particular account of first personal knowledge of mental states. On this account, first personal knowledge rests on evidence that is constituted by the presentation of mental states as something like sensations or images (or else by the presentation of mental states via intermediaries that are like sensations or images). Such an account, with sensation-like or image-like states presented to the subject as objects standing in need of interpretation, is disastrous for the idea of first person authority.

It is arguable that we do have to allow that there is something that it is like to have the (conscious, occurrent) belief that water is wet, and that this fact may, in at least some cases, figure in a personal level explanation of how a subject comes to be in a position to make the second-order judgement that he or she believes that water is wet. We might

express the point by saying that there is a phenomenology of (conscious) thought. But it would then be important to separate the idea of phenomenology or phenomenal consciousness from the idea of sensational (non-representational) properties of mental states.

It will be helpful to have before us an example of an account of first personal knowledge about beliefs that is clearly pitched at the personal level (rather than the level of subpersonal mechanisms), but that does not appeal to sensational, or image-like, presentations of conscious thoughts. Within the framework developed by Christopher Peacocke (1992), first personal knowledge of beliefs is a fairly immediate product of mastery of the concept of belief. Peacocke's proposal is that part of what is required for mastery of the concept of belief (part of the 'possession condition' for the concept of belief) is precisely that, whenever the subject has the (conscious) belief that *p*, he or she should find 'primitively compelling' the first person thought 'I believe that *p*'; and, furthermore, should find it compelling because he or she has that first-order belief. The resulting judgements constitute knowledge, Peacocke suggests, because there is a general principle linking possession conditions and knowledge (1992, p. 157):

Take any mental state of the thinker [here, believing that *p*] that a possession condition for a concept [here, the possession condition for the concept of belief] says is sufficient for the thinker finding primitively compelling a given content [here, the content: I believe that *p*] containing the concept. Then when the thinker judges that content and for the reason that he is in that state, his judgement constitutes knowledge.

To put it very roughly, the idea here is that, if a thinker makes just the judgement that is required in order to meet the conditions for the most basic mastery of a concept, then that judgement should (provided that there is nothing abnormal going on) yield knowledge.

There are, of course, many more questions that can be raised about the nature of our entitlement to self-knowledge (see Burge, 1996; Peacocke, 1996). But it seems reasonable to assume – as McLaughlin and Tye (this volume) also argue compellingly – that externalism about the contents of thoughts does not pose any insuperable achievement problem for first person authority. And, of course, it is the presumed solubility of the achievement problem for first person authority given externalism that makes the consequence problem pressing.

2. Architecturalism

We have spent some considerable time setting up the epistemological problem that is posed by philosophical arguments for externalism, when these are put together with a thesis of first person authority. We now turn, much more briefly, to the apparently parallel problem that is posed by architecturalist arguments.

2.1 Architecturalist Entailments

We have said that architecturalism about some mental property, *M*, is the thesis that whether a person (or other physical being) has *M* depends, not only on the person's (actual and counterfactual) patterns of behaviour, but also on the person's internal cognitive architecture. It may be helpful, now, to consider two examples of architecturalist arguments.

In earlier work (Davies, 1991a, 1992a), I have offered an argument with a broadly a priori character in support of the language of thought hypothesis. That argument comes in

two main stages. First, a certain kind of causal systematicity of inferential transitions is argued to be an a priori requirement for being a thinking subject. Second, it is argued that this kind of causal systematicity requires some syntactic articulation in the states that are the inputs to those inferential transitions. In short, the principal claim is that certain internal architectural conditions of systematicity and syntax – of rules and representations – are a priori requirements for being a thinking subject.

Whatever the merits of this argument concerning the language of thought, it is structurally analogous to a more familiar argument offered by Ramsey, Stich and Garon (1990). Their principal claim is that the folk psychological scheme is committed to a thesis of propositional modularity, to the effect that (1990, p. 504):

propositional attitudes are *functionally discrete, semantically interpretable*, states that play a *causal role* in the production of other attitudes, and ultimately in the production of behaviour.

In each of these two cases, the aim of the argument is to establish, by more or less a priori means, an entailment of the form:

ArchEnt $(\Box x)$ (If x has beliefs then $A(x)$)

where ‘ $A(x)$ ’ is some statement about x ’s internal cognitive architecture. Ramsey, Stich and Garon’s instantiation of the form, for example, would be:

$(\Box x)$ (If x has beliefs then x ’s internal cognitive architecture exhibits propositional modularity).

The instantiation of the form to which my own earlier work leads would be:

$(\Box x)$ (If x has beliefs then x ’s internal cognitive architecture is a language of thought architecture).

In each case, the consequent (‘ $A(x)$ ’) is a highly non-trivial claim that far outruns anything that is guaranteed by facts about x ’s patterns of behaviour. Whether this claim is true of any particular physical being cannot be settled without substantial empirical investigation.

2.2 Architecturalist Entailments and Eliminativism

An architecturalist entailment imposes a necessary condition for a physical being to be a thinking, believing subject, and so presents the possibility of an eliminativist *modus tollens*. Given any physical being whose behaviour *prima facie* warrants the attribution to it of beliefs and other attitudes, in accordance with the intentional stance (Dennett, 1987), it is a genuine epistemic possibility that the being does not meet the imposed condition on internal cognitive architecture. So a quite general question that architecturalist entailments raise is whether it is philosophically acceptable that an a priori argument should open up the epistemic possibility that we ourselves might not be thinking, believing subjects.

Suppose, for a moment, that we were to find ourselves in a disobliging world. Suppose, that is, that developments in the scientific investigation of the mind – whether in cognitive psychology or in neuroscience – were to show that the purportedly necessary conditions for us to have beliefs were not, in fact, met. Then we would have a number of theoretical options. One option would be to conclude that the substantive pieces of philosophical theory deployed in the architecturalist argument in question are wrong; that they do not provide a good elaboration and precisification of our current conception of thinking, believing subjects. Another option – the opposite extreme –

would be to abandon wholesale our folk psychological practice of describing, interpreting, and explaining people's behaviour as made up of actions performed for reasons based upon beliefs, wants, hopes, fears, and the rest.

When just these two theoretical options are in view, the thought that the second option is not genuinely available to us – that our engagement in ordinary folk psychological practice is philosophically non-negotiable – may provide a powerful source of resistance against all architecturalist arguments. But, in fact, it is not really clear how one could argue from the non-negotiability of our engagement in folk psychological practice to the incorrectness of all architecturalist analyses of the concept of being a thinking, believing subject.

One problem with the putative line of argument is that it seems to overlook the possibility that we might maintain the practice even though many of the claims made in folk psychological descriptions, interpretations, and explanations were false. A second problem is that the competing piece of conceptual analysis that is suggested by this line of argument is itself arguably out of line with our intuitive judgements about which physical beings have beliefs. What the blanket rejection of all architecturalist arguments suggests is that an analysis of the concept of a thinking, believing subject should impose no necessary conditions at all on internal cognitive architecture and, indeed, no necessary conditions that go beyond behaviour. This entails a form of behaviourism – not analytical behaviourism, to be sure, but a doctrine that we can call *supervenient behaviourism*. If two physical beings are behavioural (or, perhaps better: trajectorial) duplicates in actual and nearby counterfactual situations, then either both have beliefs, or else neither does. But this doctrine is revealed as being out of line with our intuitions when we consider imaginary examples of beings that produce the right behaviour by way of unusual internal architectures, such as the string-searching machine of Block (1981) or the Martian marionette of Peacocke (1983). The string-searching machine, which stores a finite but massive collection of interpretable sequences of behaviour, can *ex hypothesi* meet any behavioural requirements for thinking. But, as Block remarks in a later paper (1990, p. 252), 'it has the intelligence of a jukebox'.

In between the two options of, on the one hand, rejecting architecturalist arguments and, on the other hand, abandoning our folk psychological practice, there lies the possibility of conceptual negotiation. Suppose that the philosophical theory deployed in a particular architecturalist argument does correctly elaborate and precisify our current conception, and that the argument does correctly uncover the commitments of that conception. Suppose, too, that those internal architectural commitments are not met in the real world. Then it follows that we ourselves do not fall under (the best elaboration and precisification of) our current concept of a physical being with beliefs. But we may still be able rationally to sustain the greater part of our folk psychological practice if we negotiate our way to a new, revised, conception of what it is to be a thinking, believing subject. The details of the negotiations would depend on the particular discoveries that might be made in cognitive psychology and neuroscience – and also, of course, on the philosophical theories that connect those discoveries with our current conception of folk psychological phenomena.

From this suggestion, a certain picture of philosophical theorising itself emerges. The picture involves three phases. First, there is relatively pure a priori conceptual analysis. Then, there is the empirical discovery that we live in a disobliging world. Finally, there is the phase of conceptual negotiation, in which philosophical theory has to be much more answerable to the details of how the world turns out to be.

We have been considering the question of whether it is philosophically acceptable that an a priori argument should render it epistemically possible that we ourselves do not have beliefs – that we do not fall under our own current conception of thinking, believing subjects. The provisional answer that we have reached is that this is acceptable – that it is no objection to an architecturalist argument that it presents the possibility of an eliminativist *modus tollens*.

However, it might be replied that there is something unsatisfactory in this discussion of the non-negotiability of folk psychological practice. My strategy has been to suggest that, if it is to be guaranteed that those who engage in the practice really are thinking, believing subjects, then what is required is a counter-intuitive doctrine of supervenient behaviourism. But, it might be said, there is something that can be known, at least in the case of the first person singular, which goes beyond behaviour, yet has nothing to do with internal cognitive architecture. So, it would be possible to reject architecturalism without commitment to supervenient behaviourism by saying that an analysis of the concept of a thinking, believing subject should impose no necessary conditions that go beyond behaviour plus whatever can be known to be true by introspection.

This reply will not protect the status of all those who engage in folk psychological practice, since it is restricted to the first person singular. But even setting aside that point, we can note that the reply is liable to seem question-begging. For, according to architecturalism, in a disoblighing world I would not be a being with beliefs; so I could not know myself, whether by introspection or any other way, to be a being with beliefs.

The reply may be dialectically inadequate, because question-begging; but it is difficult not to sense that the appeal to first personal knowledge points to a problem with architecturalist arguments. The claim of this paper is that the real problem for architecturalism is parallel to the consequence problem for externalism given first person authority.

2.3 Architecturalist Entailments and First Person Authority

The arguments of the **(MC)** form that concern us here make use (at premise (2)) of architecturalist entailments of the form:

ArchEnt $(\Box x)$ (If x has beliefs then $A(x)$)

where ‘ $A(x)$ ’ is some statement about x ’s internal cognitive architecture. The specific form of the problematic arguments is **Arch(MC)**:

Arch(1) I have beliefs.

Arch(2) If I have beliefs then $A(\text{me})$.

Therefore:

Arch(3) $A(\text{me})$.

I do not need to engage in any detailed empirical investigation of my internal cognitive architecture to know premise Arch(1); I know with first person authority that I have many beliefs, such as the belief that water is wet. Nor, do I need to engage in empirical research in order to know premise Arch(2); according to architecturalism it is a piece of conceptual analysis. It is obvious that the conclusion Arch(3) follows from these two premises. Yet, even supposing that the conclusion is true, it is massively implausible that its truth can be known from the armchair. Questions about internal cognitive architecture cannot be settled without major programmes of empirical research. A thoroughgoing defence of architecturalist arguments requires some principled limitation on the transfer of epistemic warrant from Arch(1) and Arch(2) to Arch(3).

We have now described the epistemological problem that is posed by philosophical arguments for externalism, or for architecturalism, when these arguments are put together with a thesis of first person authority. What is needed, if we are to provide a solution to this problem, is some principled limitation on knowledge by inference. The remainder of this paper sketches some initial, tentative and provisional steps in the direction of such a solution.

3. *Scaffolding and Scepticism*

We can begin our journey towards a limitation principle on knowledge by inference by recalling an example from Wittgenstein (1969), *On Certainty*.

3.1 *Justification and Comfortable Certainty*

At *OC*, 208-211, we find this:

208. I have a telephone conversation with New York. My friend tells me that his young trees have buds of such and such a kind. I am now convinced that his tree is Am I also convinced that the earth exists?

209. The existence of the earth is rather part of the whole picture which forms the starting-point of belief for me.

210. Does my telephone call to New York strengthen my conviction that the earth exists?

Much seems to be fixed, and it is removed from the traffic. It is so to speak shunted onto an unused siding.

211. Now it gives our way of looking at things, and our researches, their form. Perhaps it was once disputed. But perhaps, for unthinkable ages, it has belonged to the *scaffolding* of our thoughts. . . .

The argument that we need to consider here is **Tree**:

Tree(1) My friend in New York has a . . . tree in his garden.

Tree(2) If my friend in New York has a . . . tree in his garden, then the earth exists.

Therefore:

Tree(3) The earth exists.

It is not obvious that Wittgenstein is arguing for a failure of closure of knowledge under known entailment. Rather, it seems that he allows that there is no harm in saying that I do know that the earth exists. But the mistake would be to think that, in order to know Tree(3), I need a justification. Thus, *OC*, 357-9:

357. One might say: “ ‘I know’ expresses *comfortable* certainty, not the certainty that is still struggling.”

358. Now I would like to regard this certainty, not as something akin to hastiness or superficiality, but as a form of life. . . .

359. But that means I want to conceive it as something that lies beyond being justified; as it were, as something animal.

Someone might respond at this point that we often find ourselves with things that we do not need. So, even if I do not need a justification for Tree(3), still the argument from Tree(1) and Tree(2) to Tree(3) does furnish a kind of justification. In order to rebut that suggestion, we can consider what Wittgenstein says in a related case (*OC*, 188):

188. It strikes me as if someone who doubts the existence of the earth at that time [150 years ago] is impugning the nature of all historical evidence. . . .

An argument to which this remark could be applied would be **Tree'**:

Tree'(1) There was a . . . tree growing in such-and-such a garden 150 years ago.

Tree'(2) If there was a . . . tree growing in such-and-such a garden 150 years ago, then the earth existed at that time.

Therefore:

Tree'(3) The earth existed 150 years ago.

Wittgenstein's thought seems to be that if Tree'(3) were subject to doubt then we could not take what would ordinarily be reckoned as historical evidence for Tree'(1) to be evidence. So, the argument from Tree'(1) and Tree'(2) to Tree'(3) can scarcely furnish a justification for Tree'(3).

What we have from Wittgenstein, then, is a form of argument that proceeds from some workaday proposition, via a readily known entailment, to a proposition that belongs to 'the scaffolding of our thoughts'. The idea to which the examples tend is that justification (or epistemic warrant) cannot be transferred from the workaday proposition via the known entailment to the scaffolding proposition. Thus, in the original example (**Tree**), I do have a justification for believing:

Tree(1) My friend in New York has a . . . tree in his garden

– a justification that is furnished by what my friend has told me about the buds, and so on, in our telephone conversation. And I can readily know:

Tree(2) If my friend in New York has a . . . tree in his garden, then the earth exists.

But that does not result in my having a justification for believing the conclusion:

Tree(3) The earth exists.

Nor, according to Wittgenstein, do I need any such justification.

3.2 *Scepticism and Cognitive Achievement*

In his 1985 British Academy lecture, 'Facts and Certainty', Crispin Wright suggests that, in general, reasonable or justified belief is transmitted over known entailment, but that this is not so when the entailed proposition falls outside the domain of what he calls *cognitive achievement* (1985, p. 470). This is not necessarily a matter of denying that knowledge is closed under known entailment. We can say that we do know the negation of the sceptical hypothesis, provided that we allow knowledge to include both justified belief and comfortable certainty. The issue is, rather, one about transfer of justification or epistemic warrant.

What our brief discussion of *On Certainty* suggests is that justification or warrant cannot be transferred to the entailed proposition because that proposition is already operating as a kind of presupposition of the justificatory, or epistemic, project. It is 'giving our researches their shape' and underpinning the relevant evidential relations.

Wright explicitly considers the issue of a limitation on transfer of epistemic warrant in the context of a sceptical response to Moore's (1959) anti-sceptical argument for the existence of the external world:

Moore(1) This is a hand before me.

Moore(2) If this is a hand before me then there is an external world.

Therefore:

Moore(3) There is an external world.

The sceptic begins by allowing that we can, indeed, know Moore(2). But still, we cannot arrive at knowledge of Moore(3) by inference from Moore(1) in the case where that latter piece of knowledge is warranted by sensory experience. For the sensory experience supports Moore(1) only in the context of our prior commitment to Moore(3). As Wittgenstein himself might have said (in the style of *OC*, 188): 'It strikes me as if someone who doubts the existence of the external world is impugning the nature of all sensory evidence.' Or, as Wright puts it (1985, p. 437):

Once the hypothesis is seriously entertained that it is as likely as not, for all I know, that there is no material world as ordinarily conceived, my experience will lose all tendency to corroborate the particular propositions about the material world which I normally take to be certain.

Wright goes on to point out (p. 458) that, taken not merely as a rejoinder to Moore's anti-sceptical argument, but as a purported sceptical argument in its own right, the force of the challenge would depend on the assumption that we need to have some evidential or inferential warrant for our commitment to Moore(3). That assumption is, of course, just what Wittgenstein denies. The scaffolding proposition, Moore(3), falls outside the domain of cognitive achievement; we accept it without reason or warrant. Then, within the framework of the scaffolding provided by Moore(3), we have the evidential support for the propositions that do need it – workaday propositions such as Moore(1).

3.3 A Limitation Principle

What I want to take from this brief discussion of Wittgenstein and Wright is a rough first shot at a principle for limiting the transfer of epistemic warrant (or limiting the scope of knowledge by inference) from premise(s) A to conclusion B.

First Limitation Principle:

Epistemic warrant cannot be transferred from A to B, even given an a priori known entailment from A to B, if the truth of B is a precondition of our warrant for A counting as a warrant.

Given the discussion from which this principle has emerged, we should expect that it can be used in discussions of scepticism in rather the same way that the denial of the closure of knowledge under known entailment is used by Robert Nozick (1981)

As is very well known, the denial of closure permits a strategy for neutralising the admission that we do not know the negation of some sceptical hypothesis. Thus, consider some workaday proposition, such as the proposition that I am working in an office in Oxford; some sceptical hypothesis, such as the hypothesis that I am a brain in a vat on Alpha Centauri; and the following argument **Scep**:

Scep(1) I am working in an office in Oxford.

Scep(2) If I am working in an office in Oxford then I am not a brain in a vat on Alpha Centauri.

Therefore:

Scep(3) I am not a brain in a vat on Alpha Centauri.

I can surely know Scep(2). If knowledge is closed under known entailment then the admission that I do not know Scep(3) involves the admission that I do not know Scep(1). But denying closure opens up the possibility of claiming that I can and do know Scep(1) and a host of other workaday propositions, even though I do not know Scep(3) – and, indeed, do not know the negation of any radical sceptical hypothesis. (I cannot, of course, know Scep(1) with first person authority; my knowledge of Scep(1) is based on empirical evidence furnished by the senses.)

We can now use the First Limitation Principle to explain why I cannot know Scep(3) by inference even though I do know Scep(1) and I know (near enough to a priori) Scep(2), that is, I know that Scep(1) entails Scep(3). The application of the limitation principle is triggered if the truth of Scep(3) is a precondition of my warrant for Scep(1) counting as a warrant. And the basic idea – in line with what Wright said about Moore's argument – is that, if I seriously entertain the brain-in-a-vat hypothesis as something that might, for all I know, be true, then I can scarcely count my sensory experience as supporting my belief that I am working in an office in Oxford.

It would be entirely reasonable for someone to complain that the First Limitation Principle, as presently formulated, is imprecise and is almost certainly open to counterexamples. A legitimate focus for dissatisfaction is the notion of a precondition; and if this is glossed simply as a necessary condition then plausible counterexamples are easily constructed. Thus, for example, consider the simple inference from the premise:

I have a warrant for p

to the conclusion:

I have a warrant for something.

It is plausible that, if I know this premise, then I can gain knowledge of the conclusion by inference. But, of course, the truth of the conclusion is a necessary condition for my warrant for the premise counting as a warrant; so application of the principle would be triggered, and transfer of warrant from premise to conclusion would be blocked.

In order to improve on the present formulation of the First Limitation Principle, the first step would be to provide a more explicit account of the relevant notion of a precondition, so as to make clear that not all necessary conditions are preconditions. But we shall not start on that here since the First Limitation Principle cannot, in any case, be used to deal with the problems of epistemic warrant that are posed by externalist and architecturalist arguments when these are put together with a thesis of first person authority. (See below, at the end of Section 4.1, for some suggestions as to how the notion of a precondition might be refined.)

In the case of externalist arguments, for example, it is difficult even to make sense of the idea that if I seriously entertain the idea that perhaps I have not had such-and-such

encounters with water then that undermines my taking as evidence what I usually take to be evidence in support of my belief that I believe that water is wet. The epistemological problems posed by arguments of the forms **Ext(MC)** and **Arch(MC)** turn on first personal knowledge of premise (1), and first personal knowledge does not seem to rest on any special kind of evidence. If we are to deal with these problems, then some other restriction on knowledge by inference is required.

4. A Second Limitation Principle and First Person Authority

The motivation that we have offered on behalf of the First Limitation Principle may seem to involve a particular picture of what our epistemic situation would be if we were brains in vats on Alpha Centauri. That picture has it, in part, that if I really were a brain in a vat on Alpha Centauri, then I would be able to frame the proposition that I am working in an office in Oxford – just as I can frame it now – but my sensory experience would no longer constitute any warrant for that belief. But this aspect of the picture of our other-worldly epistemic situation is something that many philosophers – certainly many externalist philosophers, and most notably Hilary Putnam (1981) – would reject. According to the alternative view that these philosophers would defend, the truth of the brain-in-a-vat hypothesis would actually undermine my ability to believe the proposition – indeed, even my ability to frame the hypothesis – that I am working in an office in Oxford.

4.1 A Second Limitation Principle

We may suppose that these philosophers would want to block the transfer of epistemic warrant from premises to conclusion in the argument **Scep**:

Scep(1) I am working in an office in Oxford.

Scep(2) If I am working in an office in Oxford then I am not a brain in a vat on Alpha Centauri.

Therefore:

Scep(3) I am not a brain in a vat on Alpha Centauri.

(Even someone who thinks that Scep(3) can be established by way of a transcendental argument is unlikely to think that knowledge of Scep(3) can be attained simply by inference from Scep(1) and Scep(2).) But, on their view, the truth of Scep(3) is not so much a precondition of our warrant for Scep(1) counting as a warrant as it is a precondition of Scep(1) even being a potential candidate for epistemic warrant. The truth of Scep(3) is a precondition of Scep(1) being believed or even entertained as an hypothesis. So we can obtain a limitation principle that will block the transfer of warrant from premises to conclusion in **Scep** if we simply plug this in as the triggering condition (in place of the condition that the truth of B is a precondition of our warrant for A counting as a warrant). Thus we arrive at a rough first shot at a second principle for limiting the transfer of epistemic warrant (or limiting the scope of knowledge by inference) from premise(s) A to conclusion B.

Second Limitation Principle:

Epistemic warrant cannot be transferred from A to B, even given an a priori known entailment from A to B, if the truth of B is a precondition of the knower even being able to believe the proposition A.

It would be fair to comment that the motivation that we have provided for the Second Limitation Principle has been somewhat oblique. We have simply transposed the First Limitation Principle so as to achieve the same blocking effect, while taking account of the more radical view of the cognitive and epistemic consequences of being a brain-in-a-vat. So the route to the Second Limitation Principle is bound to appear opportunistic and even ad hoc.

The intuitive idea behind both limitation principles is something like this. In any given epistemic project, some propositions will have a presuppositional status. Suppose that the focus of the project P is the proposition A, and that the investigation is carried out using method N. Then within P it is presupposed, for example, that A is a hypothesis that can be coherently entertained (can be believed, doubted, confirmed, disconfirmed); and it is also presupposed that N is a method that can yield knowledge, at least with respect to A. Suppose that B is some proposition that has this presuppositional status in project P. Then P cannot itself yield knowledge that B; nor can P play an essential role in yielding knowledge that B.

The First Limitation Principle can be regarded as an attempt at codifying this idea as it relates to the presupposition about the method: our warrant for A counting as a warrant stands in for the investigative method being such as to yield knowledge. The Second Limitation Principle likewise attempts to codify the idea as it relates to the presupposition about the hypothesis: the knower being able to believe the proposition A stands in for the hypothesis being such as can be entertained coherently. If in the principles we gloss the notion of a precondition as a necessary condition, then this amounts to saying that if B has presuppositional status in P and B is sufficient for C, then C also has presuppositional status in P.

Just as, with this gloss on the notion of a precondition, it is easy to construct counterexamples to the First Limitation Principle, so too there are counterexamples to the Second Limitation Principle (when the same gloss is used). Consider, for example, the simple inference from the premise:

I believe that water is wet

to the conclusion:

Someone is able to believe something.

It is not obvious that we should want to block the possibility of knowledge by inference here. But, the truth of the conclusion is a necessary condition for my being able to believe the premise; so application of the Second Limitation Principle would be triggered.

We need to tighten up the Second Limitation Principle (and the First Limitation Principle, too) by introducing a more constrained notion of a precondition. Rather than allowing all necessary conditions to count as preconditions, we shall, perhaps, reckon as preconditions only those necessary conditions that can be shown to be such by way of theoretical considerations that have particular salience. Salient considerations might include elements of a theory about which methods are knowledge-yielding (a theory about epistemology) and elements of a theory about which hypotheses are coherently entertainable (a theory about the contents of thoughts). Furthermore, if proposition B has presuppositional status in epistemic project P, and C can be shown to follow from B by using considerations that are already being deployed in project P, then perhaps that is enough to confer presuppositional status on C as well.

We cannot now pursue the question of whether refinements like these will yield a version of the Second Limitation Principle that is not open to obvious counterexamples, or whether some quite different line of improvement will need to be considered. Instead, we turn to confirming that the principle does indeed deal with the problems of epistemic

warrant that are posed by externalist and architecturalist arguments when these are put together with a thesis of first person authority

4.2 Externalist Arguments and First Person Authority

The epistemological problem posed by externalist arguments is that, when they are combined with a thesis of first person authority, they seem to offer me an unacceptably ‘non-empirical’ route to knowledge of substantive empirical facts about the world outside my skin.

The Second Limitation Principle deals with the problem by blocking the transfer of epistemic warrant from premises to conclusion in arguments of the form **Ext(MC)**:

Ext(1) I believe that p.

Ext(2) If I believe that p then E(me).

Therefore:

Ext(3) E(me).

The philosophical theory that issues in the externalist entailment, for example:

(\Box x) (If x believes that water is wet

then x has had such-and-such encounters with water)

and thus provides the epistemic warrant for:

If I believe that water is wet then I have had such-and-such encounters with water also has the consequence that I need to have had those encounters even to believe that I believe that water is wet. For that second-order belief is also a belief involving the concept *water*. So, if the externalist philosophical theory is correct then, according to the Second Limitation Principle, epistemic warrant cannot be transferred from Ext(1) to Ext(3), even given the a priori known entailment in Ext(2). I cannot come to know the conclusion Ext(3) by inference from the premises Ext(1) and Ext(2).

4.3 Architecturalist Arguments and First Person Authority

The epistemological problem posed by architecturalist arguments is that, when they are combined with a thesis of first person authority, they seem to offer me an unacceptably ‘non-empirical’ route to knowledge of substantive empirical facts about the world inside my skull. As would be expected given the apparent parallel between the two cases, the Second Limitation Principle deals with this problem by blocking the transfer of epistemic warrant from premises to conclusion in arguments of the form **Arch(MC)**:

Arch(1) I have beliefs.

Arch(2) If I have beliefs then A(me).

Therefore:

Arch(3) A(me).

The philosophical theory that issues in the architecturalist entailment, for example:

(\Box x) (If x has beliefs then x’s internal cognitive architecture is a language of thought architecture)

also has the consequence that my cognitive architecture must be a language of thought architecture if I am even to believe that I have beliefs. This architecturalist philosophical theory might be wrong, of course. But if it is right, then it triggers application of the Second Limitation Principle, and epistemic warrant cannot be transferred from Arch(1) to

Arch(3), even given the a priori known entailment in Arch(2). Given the Second Limitation Principle, the piece of philosophical theory that gives rise to the epistemological problem also provides for its solution. So, given that principle, it would be wrong to press the epistemological problem as an objection against architecturalist arguments.

5. *To Be Continued . . .*

If we accept the Second Limitation Principle, then we can avoid the epistemological problem that is otherwise posed by philosophical arguments for externalism about the contents of thoughts, and by philosophical arguments for architecturalism about thinking, when those arguments are put together with a thesis of first person authority. But we can scarcely be content with the present situation.

5.1 *Improving the Principle*

The Second Limitation Principle allows us to block the unacceptably ‘non-empirical’ route to knowledge of empirical facts. But we have provided little enough motivation for the principle, and we have acknowledged that, as presently formulated (and with the notion of a precondition glossed as a necessary condition), the principle is open to counterexamples.

On the other hand, we can give the principle some plausibility in terms of the idea of presuppositions in an epistemic project. And we have been able to make some suggestions as to how the principle might be improved. The question that remains, of course, is whether, as the principle is tightened up, it can continue to do the work for which it was introduced. It might turn out, for example, that despite the apparent parallel between the externalist and architecturalist cases an improved version of the Second Limitation Principle will deal with one problem and not with the other. These are issues for future work.

5.2 *Closure and Context Shifting*

The limitation principles block the transfer of epistemic warrant, and thereby limit the scope of knowledge by inference. And, as I mentioned near the outset, it is reasonable to suppose (even though this does not strictly follow) that the limitation principles generate failures of the closure of knowledge under known entailment. In a recent discussion of the way that Fred Dretske (1970) uses the denial of closure in a response to sceptical arguments (cf. Nozick, 1981), David Lewis says (1996, p. 564):

What Dretske says is close to right, but not quite. Knowledge *is* closed under implication. Knowing that I have hands *does* imply knowing that I’m not handless and deceived. Implication preserves truth – that is, it preserves truth in any given, fixed context. But if we switch contexts midway, all bets are off. . . . [I]n the sceptical argument the context switched midway, and the semantic value of the context-dependent word ‘know’ switched with it.

Lewis’s definition of knowledge is this (p. 551): ‘Subject S *knows* proposition P iff . . . S’s evidence eliminates every possibility in which not-P.’ But, as he explains, the definition ‘requires a *sotto voce* proviso’ (p. 554): ‘S *knows* that P iff S’s evidence eliminates every possibility in which not-P – Psst! – except for those possibilities that we’re properly ignoring.’ The key idea here is that in any given context certain possibilities are properly ignored. Thus (p. 564):

The premise ‘I know that I have hands’ was true in its everyday context, where the possibility of deceiving demons was properly ignored. The mention of that very possibility switched the context midway. The conclusion ‘I know that I am not handless and deceived’ was false in *its* context, because that was a context in which the possibility of deceiving demons was being mentioned, hence was not being ignored, hence was not being properly ignored. . . . If we evaluate the conclusion for truth not with respect to the context in which it was uttered, but instead with respect to the different context in which the premise was uttered, then truth is preserved.

This is not the place to explore the question of whether Lewis is right to maintain closure and Dretske is wrong to deny it. But some interest attaches to the way in which the limitation principles on transfer of epistemic warrant (or knowledge by inference) could be transposed into the context-shifting framework as constraints on which possibilities can be properly ignored (or on which possibilities remain as relevant alternatives).

Under the transposition, the Second Limitation Principle (as presently formulated) becomes a pair of constraints on which possibilities are relevant in a context.²

First Constraint

If you know that A (in context C) then (in C) all possibilities in which you are unable even to believe A are irrelevant alternatives.

Second Constraint

If you attend to the question whether B (in context D) then (in D) at least some not-B possibilities are relevant alternatives.

If the truth of B is a necessary condition for the knower even being able to believe the proposition A, then all not-B possibilities are irrelevant alternatives in context C, although at least some not-B possibilities are relevant alternatives in context D. So, given these two constraints, the condition that triggers application of the Second Limitation Principle enforces a shift of context.

Since A entails B, any not-B possibility is a not-A possibility. So, if someone’s evidence eliminates all the relevant not-A possibilities, then it eliminates all the relevant not-B possibilities. But since more not-B possibilities are relevant in context D than in context C, it may happen that someone’s evidence eliminates all the not-A possibilities that are relevant in C, but does not eliminate all the not-B possibilities that are relevant in D. In this way we get the appearance of a limitation on knowledge by inference.

Lewis suggests (p. 554): ‘Say that we *presuppose* proposition Q iff we ignore all possibilities in which not-Q.’ Then we can restate the way in which, according to the two constraints, the context shifts between premise A and conclusion B. When A is in question, we presuppose B, since B is a necessary condition for the knower even being able to believe A (First Constraint). But when B is itself in question, we cannot presuppose B, for that would trivialise the claim to know B (Second Constraint).

Here we see plainly that constraints on which possibilities are relevant in a context, like the corresponding limitation principles on knowledge by inference, can be motivated by ideas about presuppositions in epistemic projects. But further exploration of these

² I am grateful to David Lewis for discussion of these issues and for formulating the two constraints. That is not to say, of course, that he himself is committed to them.

ideas and comparison of the relative merits of the two approaches – denying closure and context shifting – must wait for another occasion.

References

- Block, N. 1981: Psychologism and Behaviorism. *Philosophical Review*, 90, pp. 5–43.
- Block, N. 1990: ‘The Computer Model of the Mind’. In D.N. Osherson and E.E. Smith (eds), *An Invitation to Cognitive Science, Volume 3: Thinking*. Cambridge, MA.: MIT Press, pp. 247–89.
- Brown, J. 1995a: ‘The Incompatibility of Anti-Individualism and Privileged Access’. *Analysis*, 55, pp. 149–56.
- Brown, J. 1995b: *Thought, the Environment and Privileged Access*. D.Phil. dissertation, University of Oxford.
- Brueckner, A. 1992: ‘What an Anti-Individualist Knows *A Priori*’. *Analysis*, 52, pp. 111–18.
- Burge, T. 1979: ‘Individualism and the Mental’. In P.A. French, T.E. Uehling and H.K. Wettstein (eds), *Midwest Studies in Philosophy, Volume 4: Studies in Metaphysics*. Minneapolis: University of Minnesota Press, pp. 73–121.
- Burge, T. 1982: ‘Other Bodies’. In A. Woodfield (ed.), *Thought and Object*. Oxford: Oxford University Press, pp. 97–120.
- Burge, T. 1986: ‘Cartesian Error and the Objectivity of Perception’. In P. Pettit and J. McDowell (eds), *Subject, Thought, and Context*. Oxford: Oxford University Press, pp. 117–36.
- Burge, T. 1988: ‘Individualism and Self-Knowledge’. *Journal of Philosophy*, 85, pp. 649–63.
- Burge, T. 1996: ‘Our Entitlement to Self-Knowledge’. *Proceedings of the Aristotelian Society*, 96, pp. 91–116.
- Davies, M. 1991a: ‘Concepts, Connectionism, and the Language of Thought’. In W. Ramsey, S. Stich and D. Rumelhart (eds.), *Philosophy and Connectionist Theory*. Hillsdale, NJ.: Lawrence Erlbaum Associates, pp. 229–57.
- Davies, M. 1991b: ‘Individualism and Perceptual Content’. *Mind*, 100, pp. 461–84.
- Davies, M. 1992a: ‘Aunty’s Own Argument for the Language of Thought’. In J. Ezquerro and J.M. Larrazabal (eds.), *Cognition, Semantics and Philosophy: Proceedings of the First International Colloquium on Cognitive Science*. Dordrecht: Kluwer Academic Publishers, pp. 235–71.
- Davies, M. 1992b: ‘Perceptual Content and Local Supervenience’. *Proceedings of the Aristotelian Society*, 92, 21–45.
- Davies, M. 1993: ‘Aims and Claims of Externalist Arguments’. In E. Villanueva (ed.) *Philosophical Issues Volume 4: Naturalism and Normativity*. Atascadero, CA.: Ridgeview Publishing Company, pp. 227–49.
- Davies, M. 1996: ‘Externalism and Experience’. In A. Clark, J. Ezquerro and J.M. Larrazabal (eds.), *Philosophy and Cognitive Science: Categories, Consciousness and Reasoning*. Dordrecht: Kluwer Academic Publishers, pp. 1–33. Reprinted in N. Block, O. Flanagan and G. Güzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*. Cambridge, MA.: MIT Press, 1997, pp. 309–28.
- Davies, M. and Humberstone, L. 1980: ‘Two Notions of Necessity’. *Philosophical Studies*, 38, pp. 1–30.
- Dennett, D.C. 1987: *The Intentional Stance*. Cambridge, MA.: MIT Press.
- DeRose, K. 1995: ‘Solving the Skeptical Problem’. *Philosophical Review*, 104, pp. 1–52.
- Dretske, F. 1970: ‘Epistemic Operators’. *Journal of Philosophy*, 67, pp. 1007–22.
- Evans, G. 1979: ‘Reference and Contingency’. *The Monist*, 62, pp. 161–89. Reprinted in *Collected Papers*. Oxford: Oxford University Press, 1985, pp. 178–213.
- Evans, G. 1982: *The Varieties of Reference*. Oxford: Oxford University Press.

- Goldman, A.I. 1993: 'The Psychology of Folk Psychology'. *Behavioral and Brain Sciences*, 16, pp. 15–28.
- Humberstone, I.L. 1996: 'Intrinsic/Extrinsic'. *Synthese*, 108, pp. 205–67.
- Jackson, F. 1994: 'Armchair Metaphysics'. In M. Michael and J. O'Leary-Hawthorne (eds), *Philosophy in Mind: The Place of Philosophy in the Study of Mind*. Dordrecht: Kluwer Academic Publishers, pp. 23–42.
- Jackson, F. 1998: *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- Jackson, F. and Pettit, P. 1993: 'Some Content is Narrow'. In J. Heil and A. Mele (eds), *Mental Causation*. Oxford: Oxford University Press, pp. 259–82.
- Khamara, E. J. 1988: 'Indiscernibles and the Absolute Theory of Space and Time'. *Studia Leibnitiana*, 20, pp. 140–59
- Langton, R. 1993: 'Stich on Intentionality and Naturalism'. In K. Neander and I. Ravenscroft (eds), *Prospects for Intentionality: Working Papers in Philosophy 3*. Canberra: Research School of Social Sciences, Australian National University, pp. 115–20.
- Langton, R. and Lewis, D. 199x: 'Defining "Intrinsic"'. *Philosophy and Phenomenological Research*, 58, pp. 333–45.
- Lewis, D. 1996: 'Elusive Knowledge'. *Australasian Journal of Philosophy*, 74, pp. 549–67.
- McDowell, J. 1984: 'De Re Senses'. *Philosophical Quarterly*, 34, pp. 283–94
- McDowell, J. 1986: 'Singular Thought and the Extent of Inner Space'. In P. Pettit and J. McDowell (eds), *Subject, Thought, and Context*. Oxford: Oxford University Press, pp. 137–68.
- McKinsey, M. 1991: 'Anti-Individualism and Privileged Access'. *Analysis*, 51, pp. 9–16.
- Moore, G.E. 1959: 'Proof of an External World'. In *Philosophical Papers*. London: Allen and Unwin, pp. 127–50.
- Nozick, R. 1981: *Philosophical Explanations*. Oxford: Oxford University Press.
- Peacocke, C. 1983: *Sense and Content*. Oxford: Oxford University Press.
- Peacocke, 1992: *A Study of Concepts*. Cambridge, MA.: MIT Press.
- Peacocke, C. 1996: 'Entitlement, Self-Knowledge and Conceptual Redeployment'. *Proceedings of the Aristotelian Society*, 96, pp. 117–58.
- Putnam, H. 1975: 'The Meaning of "Meaning"'. In *Mind, Language and Reality*. Cambridge: Cambridge University Press, pp. 215–71.
- Putnam, H. 1981: *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Ramsey, W., Stich, S. and Garon, J. 1990: 'Connectionism, Eliminativism and the Future of Folk Psychology'. In J.E. Tomberlin (ed.), *Philosophical Perspectives 4: Action Theory and Philosophy of Mind*. Atascadero, CA.: Ridgeview Publishing Company, pp. 499-533.
- Wittgenstein 1969: *On Certainty*. Oxford: Basil Blackwell.
- Wright, C. 1985: 'Facts and Certainty'. *Proceedings of the British Academy*, 71, pp. 429–72.