

Externalism, Self-Knowledge and Transmission of Warrant

Martin Davies

The conditions for thinking a certain thought must be presupposed in the thinking.¹

Externalism about some mental property, *M*, is the thesis that whether 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 that the person is embedded in that environment. The dependence here is supposed to be conceptual rather than causal; it is the kind of dependence that can be revealed by philosophical theorising. This is an armchair methodology; so, if philosophical theorising yields knowledge, then it is a kind of *armchair knowledge*. Its status as knowledge does not depend on our conducting any detailed empirical investigation of the world around us. The puzzle for discussion in this paper arises when the possibility of armchair knowledge of an *externalist dependence thesis* about mental property *M* is put together with a thesis of first-person authority for that same mental property.

For the purposes of generating our puzzle, we do not need to be concerned about a precise formulation of the notion of first-person authority. All that we need to suppose is that we each have a distinctively first-personal way of knowing that we ourselves have property *M*, when we do have it, without needing to conduct any detailed empirical investigation of the environment and our relation to it. This distinctively first-personal knowledge, self-knowledge, is another kind of armchair knowledge.

The puzzle that arises from combining externalism and self-knowledge is clearly visible when we consider the epistemic status of the premises and conclusion of arguments of the following form (Ext):

Ext(1) I am thinking that *p*.

Ext(2) If I am thinking that *p* then E(me).

Therefore:

Ext(3) E(me).

(‘E(x)’ is some statement about x’s environmental embedding.) The occurrence of the particular mental verb ‘think’ in Ext(1) is not vital; we could just as well have ‘I believe that *p*’. All that is vital is that the mental property of thinking or believing that *p* should meet two conditions. First, it should be a property that is subject to first-person authority. Second, it should be a property for which an externalist dependence thesis

¹ Burge, 1988, p. 653.

holds. Philosophical theorising should underwrite the truth of Ext(2), and we can assume that a philosophical theory will underwrite this particular claim by supporting a modal generalisation:

Necessarily(\Box x)(If x is thinking that p then $E(x)$)

By our assumptions about first-person authority and about philosophical theorising, I can have armchair knowledge of the premises Ext(1) and Ext(2). In neither case does my knowledge depend for its status as knowledge on an empirical investigation of my environment. Given armchair knowledge of those two premises, it is trivially easy to perform the *modus ponens* inference and so, it seems, to arrive at armchair knowledge of the conclusion Ext(3). But this creates a puzzle since Ext(3) is a statement about my environment and my way of being embedded in it, and we normally expect knowledge about such matters to depend on empirical investigation.

1. Transmission of Warrant and the Problem of Armchair Knowledge

In my view, the puzzle that arises from the combination of externalism and self-knowledge is not an isolated phenomenon. Rather, it is one of a number of puzzles that involve arguments of the more general 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.

Puzzles that are similar to the one under discussion here arise from philosophical theorising about the relationship between descriptions of ourselves as conscious, thinking, language-using persons and descriptions of human animals as information-processing systems (Davies, 2000). For example, there are philosophical theories according to which the cognitive architectural requirements for thought and concept deployment include the truth of the language of thought hypothesis (Davies, 1991a, 1992a). Other philosophical theories have the consequence that the objectivity of linguistic meaning requires the presence of tacit knowledge of a compositional semantic theory (Schiffer, 1993; Davies, 2000). I know from the armchair that I think many things and deploy many concepts, and I know from the armchair what many hitherto unconsidered sentences of my own language mean. But I do not have armchair knowledge about my own cognitive architecture. I cannot know from the armchair that the language of thought hypothesis is true of me; nor that my language processing system embodies tacit knowledge of a compositional semantic theory. These are matters for empirical investigation.

The general form of what I call ‘the problem of armchair knowledge’ is this. Armchair philosophical theorising supports an inference from A to B. A can be known from the armchair; B cannot be known without detailed empirical investigation of the world. So a thinker can know that A without detailed empirical investigation of the world, and can likewise know that if A then B. But it is obvious that B follows; no empirical investigation is needed to see that. So the thinker’s knowledge that A and knowledge that if A then B, and the thinker’s ability knowledgeably to draw the obvious inference, seem together to provide a route to knowledge that B – and still without any need to rise from the armchair. Yet B was supposed to be something that could only be known by way of detailed empirical investigation of the world. Knowledge that B was supposed to require an investigative, rather than an armchair, methodology.

If we are to provide a general solution to the problem of armchair knowledge, then we have to accept that, sometimes, knowing that A and knowing that if A then B, and then knowledgeably drawing the conclusion that B, does not constitute a route to knowledge that B. Sometimes, that is to say, the epistemic warrant or justification that we have for the premises of an argument is not transmitted to the conclusion, even though it is obvious that the premises entail the conclusion. The idea of a restriction on transmission of epistemic warrant may seem counter-intuitive at first. But in fact it is quite plausible that something like this is already needed for an adequate account of putative anti-sceptical arguments like G.E. Moore’s (1959; see below, section 4). In the remainder of this introductory section, I shall make four general comments by way of clarifying both the problem of armchair knowledge and the form of its proposed solution.

1. The first comment is simply that the issue is not one about validity of arguments. Arguments that seem to give rise to the problem of armchair knowledge, such as arguments of the (Ext) form, are palpably valid instances of *modus ponens*. Our problem is not logical but epistemological.
2. The second general comment is that the issue is not exactly one about closure of knowledge under known entailment. It is true that, if knowledge is not closed under known entailment, then transmission of epistemic warrant must be limited in some way. But the converse is not true in general. In a case where warrant cannot be transmitted from the premises of a valid argument to its conclusion, it might still be that anyone who knows the premises also knows, or can come to know, the conclusion. It might be, for example, that anyone who knows the premises inevitably has an independent warrant for the conclusion. Or it might be that knowledge of the conclusion does not require a warrant.

Since there can be a failure of transmission without a failure of closure, our key question is not whether someone who knows that A and knows that if A then B also knows that B. Rather, the question is whether someone who knows that A and knows

that if A then B *thereby* knows that B – whether, that is, drawing the obvious inference is a way of *achieving* knowledge of the conclusion.

3. The third general comment is that the issue is not one about confidence or subjective probability. To propose a restriction on transmission of epistemic warrant is not to suggest that rational subjects who attach a high degree of confidence to the proposition that A and the proposition that if A then B, should attach a substantially lower degree of confidence to the proposition that B.

There are, of course, cases in which it is rational to be much less confident about the conclusion of a *modus ponens* inference than about either of its premises. The conclusion of a valid argument is no less probable than the conjunction of its premises; but the probability of the conjunction of the premises may be substantially less than the probability of each of the premises taken individually. In the case of *modus ponens* arguments, this can happen if the probability of the conditional premise (if A then B) is high while the conditional probability of B given A is low. In such a case, the conditional is said to be not robust with respect to its antecedent (Jackson, 1987). But the *modus ponens* argument of the (Ext) form is not like this, and we can suppose that if the premises deserve a high degree of confidence individually then they also deserve a high degree of confidence together.

If a thinker is confident that A and confident that if A then B and also confident of their conjunction then she should be no less confident that B. In this sense, believing the premises (together) gives the thinker a reason to believe the conclusion. A thinker who believes that B under these circumstances is not subject to any criticism for doing so; she is not being doxastically reckless or irresponsible. But someone who proposes a restriction on transmission of epistemic warrant says that, even so, epistemically adequate warrants for believing the premises may not themselves add up to an epistemically adequate warrant for believing the conclusion.

4. The fourth general comment is that, in offering putative cases where warrant is not transmitted from premises to conclusion, we must take care to avoid equivocation on key notions that occur in the proposition that A, as between the first premise and the conditional premise. Thus, for example, in the case of (Ext), we must take care to avoid equivocation on the phrase ‘believe that *p*’ as between Ext(1) and Ext(2). Otherwise we invite a very natural response to the puzzle. Someone may say that it is only in a thin sense of the phrase that I have armchair knowledge that I believe that *p*, while it is only in a thick sense of the phrase that philosophical theory can support an externalist dependence thesis (Raffman, 1998).

2. Externalism about Content

I said at the outset that externalism is a thesis about dependence: whether a person has mental property M depends, at least in part, on the person’s environment and the way

that the person is embedded in that environment. It is useful to clarify the relationship between externalist dependence theses such as:

Necessarily(\Box x) (If x is thinking that p then $E(x)$)

and what I have elsewhere called constitutive and modal externalist theses (Davies, 1991b, 1992b, 1993, 1996, 1998).

Constitutive externalism (as it concerns mental property M) says that the fundamental philosophical account of what it is for an individual to have M needs to advert to the individual's physical or social environment. *Modal externalism* says that there are Twin Earth examples for M. That is, according to modal externalism, there are two possible situations w_1 and w_2 , differing in environmental conditions, such that an individual, a , has M in w_1 but a duplicate individual, b , lacks M in w_2 . According to modal externalism, having M is not a locally supervenient matter. That is, it does not depend only on what is going on inside the individual's skin.

Suppose that an externalist dependence thesis holds for mental property M:

Necessarily(\Box x) (If M(x) then $E(x)$).

Then we expect to be able to generate a Twin Earth example for M. To do so, we consider an individual, a , who has M in w_1 and therefore meets condition E in w_1 and a duplicate individual, b , who fails to meet E in w_2 and therefore lacks M in w_2 . Thus, unless it is somehow impossible to have a duplicate of a in an environment that fails to meet E, modal externalism goes along with the truth of an externalist dependence thesis. Furthermore, if an externalist dependence thesis for M is supported by philosophical theorising then it seems to be virtually guaranteed that constitutive externalism will also hold for M. So if an externalist dependence thesis holds for mental property M then it is plausible that both constitutive and modal externalism also hold for M.

If M is a constitutively, but not modally, externalist mental property, then we should not expect philosophical theorising to support an externalist dependence thesis about M. But suppose that M is modally externalist. Then it may be tempting to think that it will be a very short step from a Twin Earth example for M to a dependence thesis about M. In a Twin Earth example, a has M in w_1 and b lacks M in w_2 in virtue of a specific difference in environmental conditions between w_1 and w_2 . So, it may be said, whether any individual has M depends on environmental conditions being specifically as they are in w_1 rather than as they are in w_2 . But, in fact, it can be quite difficult to derive a specific dependence thesis from a Twin Earth example. To see this, we only have to consider the basic shape of Twin Earth examples that are used to support social externalism (Burge, 1979).

Alf is thinking that arthritis is painful, while TwinAlf is not. Alf has the concept of arthritis, though his understanding of it is incomplete; TwinAlf does not have the concept of arthritis at all. The crucial difference between Alf's social environment and TwinAlf's

relates to the linguistic practices of other people. Alf is surrounded by people who use the word ‘arthritis’ in such-and-such a way while TwinAlf is not. A convincing Twin Earth example of this kind would establish that having the concept of arthritis or thinking that arthritis is painful is not a locally supervenient matter. It would do so by highlighting the difference between meeting and not meeting a social environmental condition. But a Twin Earth example would not motivate the social externalist dependence thesis:

Necessarily(\Box x) (If x is thinking that arthritis is painful then x is surrounded by people who use the word ‘arthritis’ in such-and-such a way).

For all that the Twin Earth example shows, it may be that particular features of Alf’s internal makeup, or particular features of Alf’s non-social environment, make Alf peculiarly dependent on the standard-setting role of the linguistic community around him. If his internal constitution and his non-social environment were the same, but the surrounding community did not play that standard-setting role, then he would not count as having the concept of arthritis. In the example, TwinAlf, who shares both internal constitution and non-social environment with Alf, does indeed lack that concept. But it hardly follows that someone internally unlike Alf, or someone in a different non-social environment from Alf’s, would inevitably be without the concept of arthritis unless he enjoyed the support of a standard-setting linguistic community. Nor does it follow from the Twin Earth example that I am dependent on my linguistic community for the concept of arthritis in the way that Alf is.

Let us now consider the basic shape of Twin Earth examples that are used to support externalism about thoughts involving natural kind concepts (Putnam, 1975). Oscar is thinking that water is wet, while TwinOscar is not. Oscar has the concept of water; TwinOscar lacks that concept. The crucial difference between Oscar’s situation and TwinOscar’s is that the stuff in Oscar’s environment (water aka H₂O) is different from the stuff in TwinOscar’s environment (twater aka XYZ). A convincing Twin Earth example of this kind would establish that having the concept of water or thinking that water is wet is not a locally supervenient matter. It would do so by highlighting the difference between meeting and not meeting a physical environmental condition, namely, being surrounded by samples of water. But it would be wrong to think that a Twin Earth example would work by motivating a specific dependence thesis such as:

Necessarily(\Box x) (If x is thinking that water is wet then x is surrounded by samples of water).

For all that the example of the Oscar/TwinOscar duplicate pair shows, there could be someone else, also in TwinOscar’s waterless environment but not a duplicate of Oscar, who did have the concept of water. Perhaps there is something about the way that Oscar is internally that makes his having the concept of water peculiarly dependent on his environmental relations. All the more so, there could be someone in an environment quite

different from Oscar's or TwinOscar's but still with twater and no water, who had the concept of water. There could be, for all that the Twin Earth example shows.

The point of these last few paragraphs has been that it is not generally possible to read off an externalist dependence thesis from the difference between worlds w_1 and w_2 in a Twin Earth example. A Twin Earth thought experiment provides a counterexample to a claim of local supervenience and so licenses an existentially quantified statement about worlds and individuals. But an externalist dependence claim involves universal quantification over individuals and universal quantification over worlds. So it is unsurprising that there should be a gap between Twin Earth examples and externalist dependence claims.

3. Externalism and Self-Knowledge

We have seen that it is far from straightforward to motivate a specific externalist dependence thesis even when modal externalism is true. So consider a philosopher who holds that modal externalism is true for the mental properties of thinking thought contents that involve natural kind concepts. This externalist philosopher may still deny that externalist dependence theses about natural kind thoughts lead to the problem of armchair knowledge. Indeed, even a philosopher who accepts both modal and constitutive externalism for a wide range of mental properties could maintain that there are no true externalist dependence theses at all that give rise to the problem of armchair knowledge. I cannot show that such a position is incorrect. But my own view is that problematic externalist dependence theses cannot be avoided and that, in any case, the problem of armchair knowledge arises in many other areas of philosophical theorising. So, in the remainder of this paper, I shall pursue the question of how we might respond to the problem of armchair knowledge as it arises in the case of externalism and self-knowledge.

Let us suppose that it is possible to motivate some such dependence thesis as:

WaterDep Necessarily($\exists x$) (If x is thinking that water is wet then x is (or has been) embedded in such-and-such ways in an environment that contains samples of water).

Given that assumption, we can consider the following argument, (WaterExt), in which the conditional premise follows from WaterDep:

Water(1) I am thinking that water is wet.

Water(2) If I am thinking that water is wet then I am (or have been) embedded in an environment that contains samples of water.

Therefore:

Water(3) I am (or have been) embedded in an environment that contains samples of water.

It is worth noting two points about this example.

First, the conditional premise, Water(2), and the conclusion, Water(3), speak of an environment that contains samples of water rather than of an environment that contains samples of H₂O. The difference does not matter for the truth of the conditional premise; but it is important for the premise's epistemic status. It would be implausible that armchair philosophical theorising could deliver knowledge of a conditional premise that mentioned water in the antecedent and H₂O in the consequent, without drawing on empirical investigations to support the identity claim that water is H₂O. Speaking of samples of water rather than samples of H₂O protects the status of the conditional premise as a piece of armchair knowledge. But this does not make the conclusion something that is knowable from the armchair. I can and do know that my environment contains samples of water; but that knowledge is achieved by empirical investigation.

The second point to note is that thinking that water is wet is not to be equated with thinking something along the lines of: 'the chemical kind that exists in my actual environment and which falls from clouds, flows in rivers, is drinkable, colourless, odourless etc. is wet'. This is important if we are to avoid a charge of equivocation. (Recall the fourth general comment in section 1 above.)

It is plausible that if I were thinking a thought with the content 'the chemical kind that exists in my actual environment and which . . . etc.' then I could know in the special first-personal way that I was doing so. So the first premise, so interpreted, would be a piece of armchair knowledge. But, on this interpretation, the conditional premise is not something knowable from the philosopher's armchair; indeed, it need not even be true. Thinking a thought with that content does not require the thinker to be in an environment that contains samples of water. For a thinker can deploy in thought the description 'the chemical kind that exists in my actual environment and which . . . etc.' even though no chemical kind fits that description.

We could patch the conditional premise to make it plausibly true and knowable by way of philosophical theorising if we added a conjunct to the antecedent of the conditional and had it say:

If I am thinking that the chemical kind that exists in my actual environment and which . . . etc. is wet *and some chemical kind fits the description that I deploy in thought* then I am embedded in such-and-such ways in an environment that contains samples of water.

But then, to preserve the validity of the argument, we would need to patch the first premise as well and have it say:

I am thinking that the chemical kind that exists in my actual environment and which . . . etc. is wet *and some chemical kind fits the description that I deploy in thought*.

Of course, with the second conjunct added, this is no longer something that can be known in the special first-personal way.

In short, if thinking that water is wet were conceived as deploying a definite description in thought, then we would be open to a charge of equivocation between the first premise and the antecedent of the conditional premise. Without the equivocation, it would not be plausible that both premises could be known from the armchair. So (WaterExt) would fail as an example of the problem of armchair knowledge. If we are to use (WaterExt) as an example then we must not regard the thought that water is wet as involving a definite description ('the chemical kind . . . etc.'). It is far from obvious what is the right way to conceive of thoughts about natural kinds, but provisionally we can suppose that the thought that water is wet is about water in somewhat the same way that so-called recognition-based thoughts are about their objects (Evans, 1982; Brown, 1998). Against this presumed background, we can summarise our epistemological commentary on (WaterExt) as follows.

I can know the contents of my own thoughts in the special first-personal way; so I can have armchair knowledge of the first premise, Water(1). Also, we are supposing, externalist philosophical theorising yields armchair knowledge of the conditional premise, Water(2). But, while the conclusion, Water(3), is something that might well be known without any great difficulty, it seems to fall outside the scope of armchair knowledge; some empirical investigation is required.

The idea that there is some tension between externalism about content and first-person authority has been developed in more than one way. The problem posed by arguments like (WaterExt) might be called 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 an especially 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, in general, especially authoritative about such properties.

Approaches to the achievement problem typically make some use of the fact that the content of my second-order thought that I am thinking that water is wet, for example, embeds the content of my first-order thought that water is wet. So the content of the second-order thought is dependent on the environment in just the same way as the content of the first-order thought (Burge, 1988; Heil, 1988; Peacocke, 1999). This does not yet explain how it is that my second-order judgement amounts to knowledge. We still need a general account of how authoritative self-knowledge is possible (Burge, 1996, 1998; Peacocke, 1996, 1998, 1999). But, according to these approaches to the achievement problem, the fact about embedding can be used to show that no special

problem for the achievement of self-knowledge flows from the fact that the content of a first-order thought is, in part, externalistically determined.

An account of how self-knowledge is possible has to show how a second-order judgement that I am thinking that water is wet, made on the basis of my (first-order) thinking that water is wet, can be knowledge. According to externalism, quite independently of the epistemic status of the second-order judgement, I can make that judgement only if I have certain environmental relational properties. So, at the very starting point for an enquiry into the epistemic status of the second-order judgement, it is already guaranteed that I meet the externalist conditions for having the first-order thought. Thus, externalism poses no special problem for the achievement of self-knowledge.

In his seminal contribution to this topic, Tyler Burge says (1988, pp. 653–4):

Among the conditions that determine the contents of first-order empirical thoughts are some that can be known only by empirical means. To think of something as water, for example, one must be in some causal relation to water – or at least in some causal relation to other particular substances that enable one to theorize accurately about water. . . . To know that such conditions obtain, one must rely on empirical methods. To know that water exists, or that what one is touching is water, one cannot circumvent empirical procedures. But to *think* that water is a liquid, one need not *know* the complex conditions that must obtain if one is to think that thought.

Let us agree with this. In order to know that I am thinking that water is a liquid, I do not need to know anything of externalist philosophical theory, and I do not need to know that the conditions required by that theory actually obtain. But to the extent that the truth of externalist dependence theses is no bar to the achievement of self-knowledge, the consequence problem for first-person authority given externalism becomes pressing. Without empirical investigation, I know that I am thinking that water is a liquid. If I learn something about externalist philosophical theory then I can also know that if I am thinking that water is a liquid then certain conditions must obtain. I can draw the obvious consequence that those conditions do indeed obtain. Yet, as Burge says: ‘To know that such conditions obtain, one must rely on empirical methods’. This is the consequence problem.

Paul Boghossian presents this problem as a *reductio ad absurdum* of the combination of externalism about mental content and privileged self-knowledge. Externalism is here ‘the view that what concepts our thoughts involve may depend not only on facts that are internal to us, but on facts about our environment’ (1997, p. 161). The claim about self-knowledge is ‘that we are able to know, without the benefit of empirical investigation, what our thoughts are in our own case’ (ibid.). The conclusion for *reductio* is that, if both these claims were true, then ‘we would be in a position to know certain facts about the

world a priori, facts that no one can reasonably believe are knowable a priori' (ibid.). According to Boghossian, one or the other claim has to be rejected. My aim is to show how both claims can be maintained.

In an earlier and influential paper, Michael McKinsey presented a somewhat similar *reductio* (1981, p. 16):

[I]f you could know a priori that you are in a given mental state, and your being in that mental state conceptually or logically implies the existence of external objects, then you could know a priori that the external world exists. Since you obviously *can't* know a priori that the external world exists, you also can't know a priori that you are in the mental state in question.

But McKinsey's presentation has a number of distinctive features of which one is particularly important for our purposes.

McKinsey says that a priori knowledge is 'knowledge obtained independently of empirical investigation' (1981, p. 9). This sounds just like Boghossian, for whom the notion of a priori knowledge in play is simply that of knowledge that is available 'without the benefit of empirical investigation'. But, in fact, McKinsey makes a significant addition when he talks about the character of self-knowledge (ibid; emphasis added):

[W]e can in principle find out about these states in ourselves 'just by thinking', without launching an empirical investigation *or making any assumptions about the external physical world*.

The lesson that McKinsey draws from his *reductio* is that we do not have this special kind of knowledge of our own mental states in cases where the truth of an ascription of the mental state depends on the existence of objects external to the subject of the ascription. If the lesson is unpacked in line with McKinsey's characterisation of self-knowledge then it becomes this: We cannot know about our own externalist mental states without launching an empirical investigation or at least making some assumptions about the external physical world.

It is not obvious that this is something with which Burge would disagree. In a sentence just before the passage that I quoted a few paragraphs back, he says (1988, p. 653): 'It is uncontroversial that the conditions for thinking a certain thought must be presupposed in the thinking.' And at the end of that quoted passage, when he says, 'But to *think* that water is a liquid, one need not *know* the complex conditions that must obtain if one is to think that thought', he immediately adds: 'Such conditions need only be presupposed.' Perhaps there is a distinction to be drawn here between assumptions and presuppositions. But, on the face of it, Burge is allowing that in thinking that water is wet, or in thinking that I am thinking that water is wet, I presuppose or assume that the conditions necessary for me to think that thought do obtain. In that case, my knowledge

that I am thinking that water is wet is not knowledge that I can have ‘without making any assumptions about the external physical world’.

McKinsey uses a notion of a priori knowledge that is very strict. Suppose that, in that very strict sense, I could know a priori that I am thinking that water is wet, and could know a priori that if I am thinking that water is wet then environmental condition E holds. Then, according to McKinsey, that strict a priori warrant could be transmitted to the proposition that condition E holds. But, as McKinsey points out and as we must surely agree, it is absurd to suppose that E could be known a priori. So it cannot be that I can both know what I am thinking and know the truth of an externalist dependence thesis a priori, in that strict sense.

An externalist who also accepts a thesis of first-person authority may respond to McKinsey’s *reductio* argument by conceding one point but insisting on another. First, an externalist who accepts a thesis of first-person authority may concede that self-knowledge is not a priori in the strict sense. This is what the quotation from Burge (1988) suggests and in any case, quite apart from externalism, it is plausible that self-knowledge is not a priori in the strict sense; there seem to be empirical assumptions or presuppositions in the background (Peacocke, 1999, p. 244–5). But, unless background assumptions always stand in need of justification, this concession is consistent with the idea that self-knowledge is a priori in the weaker sense of not being justificatorily based on empirical investigation (McLaughlin, 2000; see also Field, 1996). So, second, an externalist who accepts a thesis of first-person authority may insist that both self-knowledge and knowledge of externalist dependence theses are a priori in that weaker sense.

Part of the importance of McKinsey’s *reductio* argument is that, although McKinsey himself focuses on a priori knowledge in the strict sense, the pattern of the argument can be repeated for the weaker notion of a priori knowledge (Boghossian, 1997; see also Brown, 1995). Suppose that we take a priori knowledge to be knowledge that does not rest on empirical investigation but may still depend on empirical background assumptions (such as the assumption that ‘the conditions for thinking a certain thought’ are met). This makes it much more plausible that externalist dependence theses can be known a priori and much more likely that first-person authority involves a priori self-knowledge. But it still seems absurd that I could know a priori that I am (or have been) embedded in an environment that contains samples of water.

It does not matter, of course, whether we use the term ‘a priori knowledge’ at all. Whatever label we use for knowledge that is not justificatorily based on empirical investigation, externalism and self-knowledge together give rise to a problem, an instance of the problem of armchair knowledge. The proposal of this paper is that a solution to the problem involves restrictions or limitations on the transmission of epistemic warrant.

4. Transmission of Warrant and Anti-Sceptical Arguments

Earlier (in section 1), I acknowledged that the idea of limitations on the transmission of epistemic warrant may seem counter-intuitive at first. But I suggested that some restriction is already needed for an adequate account of putative anti-sceptical arguments like Moore's:

Moore(1) Here is one hand and here is another.

Moore(2) If here is one hand and here is another then an external world exists.

Therefore:

Moore(3) An external world exists.

Of this purported anti-sceptical argument, Barry Stroud says (1984, p. 86):

Once we are familiar with the philosophical problem of our knowledge of the external world, I think we immediately feel that Moore's proof is inadequate.

The sense that what Moore offered is inadequate as an anti-sceptical argument does, indeed, seem to be compelling. We want to allow, with Moore, that the conditional premise can be known by a very elementary piece of philosophical theorising (or conceptual analysis); and we want to allow that the first premise can be known just by looking at one hand and then the other. The argument is palpably valid; but it seems wrong to suppose that our knowledge of the premises could provide us with a route to knowledge of the conclusion, or that the conclusion has epistemic warrant transmitted to it from the premises.

When I say that epistemic warrant is not transmitted from the premises to the conclusion of Moore's argument, I am not suggesting that someone who confidently believes Moore(1) and Moore(2) would be wrong to believe Moore(3). I am not even suggesting that the belief in Moore(3) is epistemologically out of order. The proposal is, rather, that the epistemically adequate warrants for believing the premises may not themselves add up to an epistemically adequate warrant for believing the conclusion.

We observed (the second general point in section 1) that it is possible to deny transmission of warrant without denying closure of knowledge under known entailment. Even if warrant is not transmitted from the premises to the conclusion of a particular valid argument, it might be that anyone who knows the premises also knows the conclusion because it is possible to have knowledge of the conclusion without any warrant at all.

Alternatively, it might be that anyone who has a warrant for the premises inevitably has some other, prior and independent, warrant for the conclusion. In principle, it seems that there are two significantly different ways in which this could happen. In one kind of case, anyone who has a warrant for the premises inevitably already has a warrant for the conclusion, and this warrant for the conclusion does not contribute to the warrant for the

premises. This pattern of warrants would be consistent with having closure of knowledge without transmission of warrant. But the pattern does not actually guarantee non-transmission, for it might be that the warrant is transmitted from premises to conclusion, but that the warrant for the conclusion is not needed.

In fact, it might be suggested that this is the situation with Moore's argument. On this view, anyone who looks at one hand and then the other and thereby achieves a perceptual warrant for Moore(1) already has a more than adequate warrant for Moore(3) provided by perception of countless other objects. So (continuing with the suggestion), the warrant for Moore(1), provided by perception of two hands, is indeed transmitted to Moore(3). But it is not needed, as the conclusion Moore(3) is already as warranted as can be. According to this suggestion, an illusion of non-transmission is created by a kind of epistemic 'ceiling effect'. It is not easy definitively to rebut this suggestion. But anyone who is convinced, with Stroud, that Moore's premise, 'Here is a hand and here is another', does not warrant the anti-sceptical conclusion is likely to retain their conviction if we add to the premise 'and another and another . . .'. What seems implausible is not just that one little piece of perceptual evidence could defeat the sceptic, but rather that any amount of perceptual evidence could perform that epistemic service.

There is a second kind of case in which anyone who has a warrant for the premises of an argument inevitably has some prior warrant for the conclusion. In this kind of case, the prior warrant for the conclusion figures as a component in the warrant for the premises. This pattern of warrants involves a kind of epistemic circularity, but it is not obvious that it constitutes non-transmission of warrant. Consider, for example, a very simple case of this kind. The premises of the argument are A and B, and the argument proceeds through the conjunction A&B to the first conjunct A. Clearly, anyone who has a warrant for the premises already has a warrant for the conclusion. But it is far from clear that the warrant for the premises is not transmitted to the conclusion. In fact, the most natural account of this case would be that the warrants for the two premises A and B add up to a warrant for the conjunction A&B and that this amounts to a warrant for the conclusion A. But the transmitted warrant for the conclusion A is no more epistemically adequate than the warrant for A&B and this, in turn, is no more epistemically adequate than the warrant for B. So the pointlessly indirect inferential route to the conclusion A results in transmission of the warrant for the premise A; but the transmitted warrant is liable to be contaminated and weakened along the way.

We began from the idea that non-transmission of epistemic warrant is consistent, in principle, with closure of knowledge under known entailment. It might be that knowledge of the conclusion does not require a warrant; or it might be that knowledge of the conclusion is provided by a prior warrant. We have considered two kinds of case in which there is a prior warrant for the conclusion. The first kind of case was consistent with, but

did not guarantee, non-transmission; the second kind of case seemed to amount to pointlessly circular transmission of warrant. So, although it remains plausible that Moore's argument is a case of non-transmission of warrant, we do not yet have any account of how this non-transmission arises.

Let us turn to the possibility that knowledge of the conclusion of Moore's argument, Moore(3), does not require a justification or warrant. In our discussion of the a priori of self-knowledge (in the previous section), we saw that there seem to be empirical background assumptions for self-knowledge. But, provided that background assumptions do not always stand in need of justification, this is consistent with the claim that self-knowledge is a priori in the sense of not being justificatorily based on empirical investigation. Now I want to focus on the role of Moore(3) as a background assumption for ordinary perceptual knowledge.

Perceptual knowledge does seem to depend on the assumption that an external world exists. Thus, for example, Crispin Wright says (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.

But we need to be explicit about the relationship between knowledge and justification or warrant. On one conception of this relationship, any proposition that figures as a background assumption in a project of warranting must itself be warranted, or at least be susceptible of being warranted. According to this first conception, unless the conclusion of Moore's argument can somehow be warranted, what we ordinarily take for perceptual knowledge lacks a proper warrant. But there is an alternative conception of the relationship between knowledge and warrant. According to this second conception, the fact that an unwarranted proposition is a background assumption in a warranting project does not always reflect negatively on the epistemic status of propositions that are warranted by the project. There are assumptions that we make without warrant or justification, but not without entitlement or right. It is a substantive task for epistemology to provide an account of the nature of this entitlement. But for present purposes I shall, without giving such an account, simply proceed on the basis of this second conception of the relationship between knowledge and warrant. Some background assumptions are epistemically in good order; and if they are epistemically in good order then we can regard them as cases of knowledge without warrant.

If this conception is correct, then we can offer Moore's argument as a case of non-transmission of epistemic warrant without being committed to saying that it is a counterexample to the closure of knowledge under known entailment. Non-transmission arises because the conclusion of the argument plays a crucial role as a background

assumption in the project of warranting the premises. It is only against the background of the assumption of Moore(3) that the perceptual warrant for Moore(1) counts as a warrant.

In an earlier paper (1998), I suggested a limitation on transmission of epistemic warrant along the following lines:

First Limitation Principle (first version):

Epistemic warrant cannot be transmitted from the premises of a valid argument to its conclusion if, for one of the premises, the truth of the conclusion is a precondition of our warrant for that premise counting as a warrant.

This principle appears to have the desired consequence that epistemic warrant cannot be transmitted from the premises to the conclusion of Moore's argument. But, in this initial formulation, the principle is problematic in a number of respects. It makes use of the unexplained notion of a precondition; and if this is interpreted simply as a necessary condition then the First Limitation Principle is certainly open to counterexamples.

In order to improve the formulation of the principle and to provide it with some clearer motivation, we need to make use of the idea that failure of transmission of epistemic warrant is the analogue, within the thought of a single subject, of the dialectical phenomenon of begging the question (McLaughlin, 2000; Wright, 2000). It is often said that Moore's argument begs the question against the sceptic, but what we need is an explicit account of what makes an argument question-begging.

Here I shall make use of Frank Jackson's account according to which an argument can be said to beg the question when 'anyone – or anyone sane – who doubted the conclusion would have background beliefs relative to which the evidence for the premises would be no evidence' (1987, p. 111). When a speaker advances an argument for a conclusion, he implicitly offers evidence or other considerations in support of the premises. These considerations support the premises against the background of particular assumptions. But the speaker's background assumptions might not be shared by the hearer to whom the argument is directed. In particular, it may be that the background assumptions relative to which the offered considerations support the premises cannot be rationally combined with doubt about the conclusion. In that case, the argument, with its premises supported by those implicitly offered considerations, 'could be of no use in convincing doubters' (ibid., p. 112).

Jackson's account of begging the question, together with the parallel between this and non-transmission of epistemic warrant, suggests a better formulation of the limitation principle.

First Limitation Principle (revised version):

Epistemic warrant cannot be transmitted from the premises of a valid argument to its conclusion if, for one of the premises, the warrant for that premise counts as a

warrant only against the background of certain assumptions and acceptance of those assumptions cannot be rationally combined with doubt about the truth of the conclusion.

In fact, we can slightly generalise Jackson's account of begging the question. An argument will be useless for convincing those who doubt the truth of the conclusion if, for one of the premises, that doubt cannot be rationally combined with acceptance of the background assumptions that are crucial if the implicitly offered considerations are to support that premise. But there are weaker conditions under which an argument will be equally useless for convincing doubters. The person advancing an argument implicitly offers considerations in support of each of the premises. For each premise, there may be crucial background assumptions. If doubt about the truth of the conclusion rationally undermines acceptance of background assumptions in such a way that not all the premises remain supported, then that is enough to render the argument useless. This could happen because doubt about the truth of the conclusion, when combined with acceptance of the considerations offered in support of all but one of the premises, rationally undermines acceptance of the background assumptions that are crucial for support of the remaining premise. This complex relationship of rational influence between doubt about the truth of the conclusion and acceptance of background assumptions may be quite difficult to discern. But it seems appropriate to extend the notion of begging the question to include cases with this more complex and less obvious structure.

Pursuing the parallel with non-transmission of epistemic warrant, we can now propose:

First Limitation Principle (generalised version):

Epistemic warrant cannot be transmitted from the premises of a valid argument to its conclusion if, for one of the premises, the warrant for that premise counts as a warrant only against the background of certain assumptions and acceptance (i) of those assumptions and (ii) of the warrants for the other premises cannot be rationally combined with doubt about the truth of the conclusion.

No doubt this principle will still face counterexamples. Certainly it could be further improved. But it combines two virtues: it is reasonably clearly motivated and it underwrites the non-transmission of epistemic warrant in Moore's argument.

I hope that this section has made the idea of limitations on the transmission of epistemic warrant less counter-intuitive. But we are still some way from a solution to the instance of the problem of armchair knowledge that arises from externalism and self-knowledge. After all, Moore's argument is not of the (Ext) form or the more general (MC) form; nor is our knowledge of the first premise of Moore's argument armchair knowledge.

Furthermore, Moore's argument is a very plausible example of begging the question, but the argument (WaterExt):

Water(1) I am thinking that water is wet.

Water(2) If I am thinking that water is wet then I am (or have been) embedded in an environment that contains samples of water.

Therefore:

Water(3) I am (or have been) embedded in an environment that contains samples of water.

is not intuitively or obviously question-begging (Warfield, 1998). So it is not clear how anything like the First Limitation Principle, which was motivated by the idea of a parallel between begging the question and non-transmission of warrant, could be used to solve the problem that arises from externalism and self-knowledge.

5. Transmission of Warrant and Externalist Arguments

In the earlier paper (1998) that I have already mentioned, I proposed a principle that was intended to apply to arguments of the (Ext) form such as (WaterExt).

Second Limitation Principle (first version):

Epistemic warrant cannot be transmitted from the premises of a valid argument to its conclusion if, for one of the premises, the truth of the conclusion is a precondition of the knower even being able to believe that premise.

This principle certainly has the desired result. According to externalist philosophical theorising, my being embedded in an environment that contains water is a necessary condition for my believing or even thinking that water is wet. It is also a necessary condition for my thinking any other thought in which the concept of water is deployed; in particular, for my thinking that I am thinking that water is wet. The very same philosophical theory that gives rise to the problem by supporting the externalist dependence thesis:

Necessarily(If I am thinking *that water is wet* then I am (or have been) embedded in such-and-such ways in an environment that contains samples of water).

also supports the thesis:

Necessarily(If I am thinking *that I am thinking that water is wet* then I am (or have been) embedded in such-and-such ways in an environment that contains samples of water).

This thesis triggers the Second Limitation Principle and if the principle is correct then the problem is solved.

However, given just this early version of the Second Limitation Principle, the situation is far from satisfactory. The worry is not, primarily, that the principle is open

to counterexamples, but rather that no independent motivation for the principle has been provided. In short, the Second Limitation Principle appears to be completely *ad hoc*.

We need to connect the Second Limitation Principle with the notion of a background assumption so that we can draw on some of the same ideas that were used to motivate the revised and generalised versions of the First Limitation Principle. In particular, as a first step, we need to make it plausible that something like the assumption that I am able to believe that I am thinking that water is wet is a background assumption for my first-personal knowledge that I am thinking that water is wet.

But surely this is plausible. For any proposition (or thought content) A, if there were no such proposition as A then nothing could amount to a justification for believing A. If there were no such thing to think as A, no such thing to believe or doubt, no such thing to confirm or disconfirm, then there could be no prospect of anything counting as an epistemically adequate warrant for A. So, the assumption that there is such a proposition for me to think as *that I am thinking that water is wet* is a background assumption for my first-personal knowledge (and, indeed, for any kind of knowledge) that I am thinking that water is wet.

We are now in a position to generate revised and generalised versions of the Second Limitation Principle. We simply adapt the revised and generalised versions of the First Limitation Principle by focusing on a particular assumption against the background of which a warrant for a premise counts as a warrant; namely, the assumption that there is such a proposition (there is such a thing to think) as that premise. The results are as follows.

Second Limitation Principle (revised version):

Epistemic warrant cannot be transmitted from the premises of a valid argument to its conclusion if, for one of the premises, acceptance of the assumption that there is such a proposition for the knower to think as that premise cannot be rationally combined with doubt about the truth of the conclusion.

Second Limitation Principle (generalised version):

Epistemic warrant cannot be transmitted from the premises of a valid argument to its conclusion if, for one of the premises, acceptance (i) of the assumption that there is such a proposition for the knower to think as that premise and (ii) of the warrants for the other premises cannot be rationally combined with doubt about the truth of the conclusion.

It remains to confirm that one or the other of these principles blocks the transmission of epistemic warrant from the premises to the conclusion of the argument (WaterExt) and so provides a solution to the instance of the problem of armchair knowledge that arises from externalism and self-knowledge.

The revised version of the principle is not adequate to this task. There is no immediately obvious incompatibility between, on the one hand, acceptance of the assumption that there is such a thing for me to think as that I am thinking that water is wet and, on the other hand, doubt as to whether I am (or have been) embedded in an environment that contains samples of water. It is only in the context of a philosophical theory of externalism that there is a tension between this acceptance and this doubt.

But the generalised version of the principle does have the consequence that epistemic warrant is not transmitted from the premises of WaterDep to the conclusion. The warrant for the conditional premise, Water(2), is a piece of philosophical theory that supports both the externalist dependence thesis:

Necessarily (If I am thinking *that water is wet* then I am (or have been) embedded in such-and-such ways in an environment that contains samples of water).

and the thesis:

Necessarily (If I am thinking *that I am thinking that water is wet* then I am (or have been) embedded in such-and-such ways in an environment that contains samples of water).

The theory supports the externalist dependence thesis because it supports the claim:

Necessarily (If there is such a thing for me to think as *that water is wet* then I am (or have been) embedded in such-and-such ways in an environment that contains samples of water).

Equally, it supports the claim:

Necessarily (If there is such a thing for me to think as *that I am thinking that water is wet* then I am (or have been) embedded in such-and-such ways in an environment that contains samples of water).

So acceptance (i) of the assumption that there is such a thing for me to think as Water(1) and (ii) of the warrant for Water(2) cannot be rationally combined with doubt about the truth of Water(3). According to the generalised version of the Second Limitation Principle, then, epistemic warrant cannot be transmitted from the premises Water(1) and Water(2) to the conclusion Water(3).

The combination of externalism and self-knowledge has the consequence that I can know both the premises of the argument from the armchair. This is problematic, because it is implausible that I can know the conclusion from the armchair. The Second Limitation Principle, in its generalised version, provides a solution to this instance of the problem of armchair knowledge.

Earlier, I distinguished between the achievement problem and the consequence problem for first-person authority given externalism. The key idea in dealing with the achievement

problem is that the content of my second-order thought that I am thinking that water is wet embeds the content of my first-order thought that water is wet. So the content of the second-order thought is dependent on the environment in just the same way as the content of the first-order thought. This has the result, we noted, that at the very starting point for an enquiry into the epistemic status of the second-order judgement, it is already guaranteed that I meet the externalist conditions for having the first-order thought; so externalism poses no special problem for the achievement of self-knowledge. This same idea, that the conditions for thinking the first-order thought are included amongst the conditions for thinking the second-order thought, also plays a vital role in triggering the application of the Second Limitation Principle and solving the consequence problem.

This is not yet to say that all the resources that we need to solve the consequence problem are present in Burge's (1988) solution to the achievement problem. But Burge also says (1988, p. 653): 'It is uncontroversial that the conditions for thinking a certain thought must be presupposed in the thinking.' If this is right then the conditions for thinking the second-order thought, and so also the conditions for thinking the first-order thought, are presupposed in the thinking of the second-order thought. The truth of the conclusion, Water(3), is presupposed in the thinking of the premise, Water(1). I have been developing the idea that this is why epistemic warrant cannot be transmitted from Water(1) to Water(3) via the externalist dependence thesis, Water(2).²

² Versions of this material were presented at a conference on externalism and self-knowledge held at the University of Bristol in February 1999 and in a symposium at the Pacific Division meeting of the American Philosophical Association held in Berkeley in April 1999. I am grateful to Antonia Barke, Helen Beebe, Paul Boghossian, Bill Brewer, Jessica Brown, Mark Greenberg, Frank Jackson, Brian Loar, Kirk Ludwig, Michael McKinsey, Brian McLaughlin, Michael Martin, Christopher Peacocke, Paul Pietroski, Sarah Sawyer, Stephen Schiffer, Ernest Sosa, Tom Stoneham and Crispin Wright for comments and conversations.

References

- Boghossian, P.A. 1997: What the externalist can know a priori. *Proceedings of the Aristotelian Society*, 97, pp. 161–75.
- Brown, J. 1995: The incompatibility of anti-individualism and privileged access. *Analysis*, 55, pp. 149–56.
- Brown, J. 1998: Natural kind terms and recognitional capacities. *Mind*, 107, pp. 275–303.
- 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. 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.
- Burge, T. 1998: Reason and the first person. In C. Wright, B.C. Smith and C. Macdonald (eds), *Knowing Our Own Minds*. Oxford: Oxford University Press, pp. 243–70.
- 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, pp. 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. 1998: Externalism, architecturalism, and epistemic warrant. In C. Wright, B.C. Smith and C. Macdonald (eds), *Knowing Our Own Minds*. Oxford: Oxford University Press, pp. 321–61.
- Davies, M. 2000: Conscious thinking persons and their information-processing underpinnings. *Philosophical Explorations*, 3, pp. 43–62.

- Evans, G. 1982: *The Varieties of Reference*. Oxford: Oxford University Press.
- Field, H. 1996: The apriority of logic. *Proceedings of the Aristotelian Society*, 96, pp. 359–79.
- Heil, J. 1988: Privileged access. *Mind*, 97, pp. 238–51.
- Jackson, F.C. 1987: *Conditionals*. Oxford: Blackwell Publishers.
- McKinsey, M. 1991: Anti-Individualism and privileged access. *Analysis*, 51, pp. 9–16.
- McLaughlin, B.P. 2000: Self-knowledge, externalism and skepticism. *Proceedings of the Aristotelian Society Supplementary Volume* 74, pp. 93–117.
- Moore, G.E. 1959: Proof of an external world. In *Philosophical Papers*. London: Allen and Unwin, pp. 127–50.
- Peacocke, C. 1996: Entitlement, self-knowledge and conceptual redeployment. *Proceedings of the Aristotelian Society*, 96, pp. 117–58.
- Peacocke, C. 1998: Conscious attitudes, attention, and self-knowledge. In C. Wright, B.C. Smith and C. Macdonald (eds), *Knowing Our Own Minds*. Oxford: Oxford University Press, pp. 63–98.
- Peacocke, C. 1999: *Being Known*. Oxford: Oxford University Press.
- Putnam, H. 1975: The meaning of ‘meaning’. In *Mind, Language and Reality*. Cambridge: Cambridge University Press, pp. 215–71.
- Raffman, D. 1998: First-person authority and the internal reality of beliefs. In C. Wright, B.C. Smith and C. Macdonald (eds), *Knowing Our Own Minds*. Oxford: Oxford University Press, pp. 363–9.
- Schiffer, S. 1993: Actual-language relations. In J.E. Tomberlin (ed.), *Philosophical Perspectives, 7: Language and Logic*. Atascadero, CA: Ridgeview Publishing Company, pp. 231–58.
- Stroud, B. 1984: *The Significance of Philosophical Scepticism*. Oxford: Oxford University Press.
- Warfield, T.A. 1998: A priori knowledge of the world: Knowing the world by knowing our minds. *Philosophical Studies*, 92, pp. 127–47.
- Wright, C. 1985: Facts and certainty. *Proceedings of the British Academy*, 71, pp. 429–72.
- Wright, C. 2000: Cogency and question-begging: Some reflections on McKinsey’s paradox and Putnam’s proof. *Philosophical Issues*, 10 (published as a supplement to *Noûs*), pp. 140–63.