

In the armchair, down and out

MARTIN DAVIES

1. Introduction

Sitting in the philosopher's armchair, I am not engaged in any detailed empirical investigation of the world. But, as I pursue philosophy's distinctive armchair methodology, I sometimes come upon arguments that appear to disclose requirements for thought. According to some of these arguments, being a thinking person requires having the right kind of history, or having the right kind of cognitive architecture. According to other arguments, being able to think about particular topics requires being a member of a community of speakers, or being in contact with the right kinds of stuff. These arguments have the potential to raise an epistemological problem. For, suppose that armchair philosophical arguments such as these can yield knowledge about requirements for thought, and suppose too that I can know from the armchair that I am a thinking being who has thoughts about various particular topics. Then I seem to have a route to armchair knowledge about my history, my cognitive architecture, my community, and my material environment – knowledge about these things that does not depend on detailed empirical investigation of the world.

In these lectures, I shall start (today) by showing how this epistemological problem – the problem of armchair knowledge – arises from arguments in a variety of areas of philosophy. Then (in the other two lectures), I shall try to provide a principled solution to the problem. I shall not focus on arguments about history or community, but on *architecturalist* arguments and *externalist* arguments. Armchair arguments of these two kinds start from everyday folk-psychological claims about our thoughts and their contents and reach either *down*, to the cognitive machinery that underpins our thinking, or *out*, to the world that our thoughts are about.

1.1 Philosophical methodology and conceptual analysis

Suppose you were to think that the methodology of philosophy does not differ *at all* from the methodology of empirical investigation of the world. Or suppose you were to think that philosophy differs from the sciences *only* in that philosophy integrates information from many different sciences. Then you would probably not expect philosophical arguments to be epistemologically problematic in the way that I have just described. But, while the prospect or risk of instances of the problem of armchair knowledge depends on there being something distinctive about philosophy, it does not depend on an especially purist conception of philosophical practice. It does not depend on the idea that philosophy's methodology is strictly or purely *a priori*, for example.

It is sometimes said that what philosophers do is *conceptual analysis*. If this is to be anything like the truth, then we must take a broad view of what conceptual analysis amounts to. It is not a matter of recording the ordinary uses of words or the unreflective judgements of the folk. It is not primarily a matter of stating necessary and sufficient conditions for the correct application of concepts. And it is not necessarily motivated by reductionist ambitions. The business of conceptual analysis must include, at least, clarification, elucidation and elaboration of the concepts that we use in our thought about the world and about ourselves.¹

Conceptual analysis has often been conceived in a narrower way. Thus, in the Introduction to *Individuals*, Peter Strawson speaks of one mid-twentieth-century conception when he says (1959, p. 9): 'Up to a point, the reliance upon a close

examination of the actual use of words is the best, and indeed the only sure, way in philosophy.’ But the problem is that conceptual analysis, thus narrowly conceived, is inadequate to the philosopher’s task of plotting the contours of our conceptual scheme. So, Strawson continues (ibid., p. 10): ‘The structure [the philosopher] seeks does not readily display itself on the surface of language, but lies submerged. He must abandon his only sure guide when the guide cannot take him as far as he wishes to go.’ The methodology of philosophy involves construction of substantive theories that far outrun what can be gathered by examining the use of words and far outrun what is truistic or platitudinous; and it involves adoption of some theories rather than others on the basis of inference to the best philosophical explanation.

Whether philosophy’s sure guide is close examination of the use of words, or cataloguing of platitudes, or something else, we try to go further in philosophy than our only sure guide can take us.²

2. Aunty’s armchair argument

In papers published some years ago (Davies, 1991a, 1992a), I offered an armchair argument for the language of thought (LOT) hypothesis or, rather, for the conditional claim that if x is a thinking being then x is an LOT being. I called it ‘Aunty’s own argument for the language of thought’, alluding to a character who sometimes appears in Jerry Fodor’s writings. He represents her as a conservative figure, more likely to favour connectionism than to accept that there are good reasons to adopt the language of thought hypothesis.³

Aunty’s armchair argument relies on a very general background assumption that the thinking being under consideration is physically embodied. It also relies on a more specific assumption that personal-level events of conscious judgement and thought are underpinned by occurrences of physical configurations belonging to kinds that figure in the science of information-processing psychology. Since these physical configurations can be assigned the contents of the thoughts that they underpin, they are ‘proposition-sized’ bearers of causal powers. This assumption is what Fodor (1985, 1987) calls *intentional realism* and it is close to Ramsey, Stich and Garon’s assumption of *propositional modularity* (1990, p. 504): ‘[P]ropositional attitudes are *functionally discrete, semantically interpretable*, states that play a *causal role* in the production of other propositional attitudes, and ultimately in the production of behavior.’

In my provisional view, we are committed to this assumption by some of our everyday practices of mental talk and explanation. I am inclined to think that philosophical theorising would support the conditional claim that if x is a thinking being then x meets the requirement of propositional modularity. If I am right about this then the assumption can be discharged. If I am wrong, then we can regard Aunty’s argument as an argument for the claim that *if* x is a thinking being *and* x meets the requirement of intentional realism or propositional modularity *then* x is an LOT being.⁴ In any case, the argument proceeds in two main steps.

The first step makes use of neo-Fregean resources. Thinking involves the deployment of concepts and having concepts involves commitments to certain forms or patterns of inference. Commitment to a particular form of inference is not just commitment to each of a number of inferences that happen to instantiate that form. Rather, the commitment is to accept or perform those inferences *in virtue of their form*. The form of the inferences should figure, somehow, in the causal explanation of the thinker’s performing those inferences.⁵ But it is not obvious what this requirement comes to. We need to unpack the idea of performing inferences in virtue of their form without requiring that a thinker must

be able to specify the form of the inferences that he or she makes. Still less should it be required that the thinker must offer an explicit account of the form of the inferences as part of his or her reason for making the inferential transitions.

My proposal, which is the result of a kind of inference to the best philosophical explanation, is that performing inferences in virtue of their form involves meeting the conditions for *tacit knowledge* of the corresponding inferential rule. The notion of tacit knowledge that is in play here goes well beyond that of being able to do something but unable to say how one does it (see Polanyi, 1967). What it is for a rule to be tacitly known can be elucidated in terms of a structure in the explanations of causal processes. Roughly, the transitions that conform to the rule should have a common causal explanation.⁶

The second step of Aunty's argument then makes use of a quite general connection between tacit knowledge of rules, elucidated in this way, and syntactically structured representations. To see the basic idea behind this connection, consider a transition-mediating mechanism that embodies tacit knowledge of a rule. This mechanism operates as a causal common factor to mediate transitions that conform to the rule; and the inputs that engage the mechanism are representations that have some semantic property in common. So, those input representations must have in common a physical property that meets two conditions. It is correlated with the semantic property that they share; and, by engaging the mechanism, it is a determinant of causal consequences. Physical properties that meet these conditions meet the minimal requirements for being *syntactic* properties of representations (Fodor, 1987, pp. 16–21).

Thus, if inferential transitions are to be explained in terms of tacit knowledge of rules of inference, then the physical configurations that underpin thoughts must have syntactic properties that encode conceptual constituents of the contents of those thoughts (and encode the structural relations amongst those constituents).⁷

In summary, the two steps of Aunty's argument take us from the initial assumptions of physical embodiment and intentional realism to the conclusion that the physical configurations that underpin thoughts are syntactically structured. Because we have used only the minimal notion of a syntactic property, this consequence may be less than is required by some more robust, empirically motivated, versions of the LOT hypothesis. But, even when the assumption of intentional realism is explicitly included in the antecedent of the conditional, the conclusion of Aunty's argument is not trivial or obvious. Many philosophers and cognitive scientists accept intentional realism but deny the LOT hypothesis. There is no doubt that, in principle, it is possible to produce input-output performance that conforms to rules of inference over large finite domains of propositionally modular inputs, without having the inputs exhibit even this minimal kind of syntactic structure.⁸

2.1 The worry about eliminativism: An intuition of non-negotiability

The very fact that the conclusion of Aunty's armchair argument is no platitude may lead some people to think that the argument is a philosophical liability. The question whether we do or do not have an LOT cognitive architecture is a substantive empirical one to be settled by detailed empirical investigation and by comparison of results from competing research programmes. It seems to be conceivable, whether likely or unlikely, that the evidence should support the hypothesis that we are not LOT beings – or even the hypothesis that we are not LOT beings but that still intentional realism is true of us. In that case, Aunty's argument would allow an eliminativist *modus tollens*. From the

empirical discovery that we are not LOT beings we would be able to conclude that we are not thinking beings.⁹

The worry here is that there is a powerful intuition to the effect that it is not genuinely open to us to abandon altogether our familiar descriptions of ourselves and others as believing and wanting things, as hoping and fearing things, as engaging in reasoning and planning. Of course, our opinions about detailed psychological matters may be wrong in many ways. But our basic commitments to everyday folk psychological description, explanation, and prediction seem to be philosophically non-negotiable. Standing ready to perform an eliminativist *modus tollens* if the empirical evidence from information-processing psychology and neuroscience turns out one way rather than another seems to be incompatible with proper respect for that intuition of non-negotiability.

The lesson that many will draw is that there must be something wrong with Aunty's argument. Indeed, the lesson seems to extend to any philosophical argument that purports to uncover requirements for thought, if the requirements are sufficiently substantive that they conceivably might not be met. But my view is that, by slightly complicating the account of our conception of a thinking being, we can respect the intuition of non-negotiability even while embracing the philosophical theories that support Aunty's argument.

2.2 Responding to the worry about eliminativism

In order to explain the complication, it is helpful to use the idea of introduction and elimination rules for a concept. These rules would figure as components in a conception. Some of these rules would be primitive or non-derived rules, which would appear in the core conception required for grasp of the concept. Other rules figuring in a conception might be derived rules, and might go far beyond anything required for grasp of the concept. Some philosophers would say that a concept is individuated by its associated non-derived introduction and elimination rules. But I do not mean to take a stand on that issue.

Suppose, for a moment, that the philosophical theories supporting Aunty's argument do provide the best clarification, elucidation and elaboration of our current conception of a thinking being. Suppose that the argument itself correctly draws out a requirement for conceptualised thought. We can regard this as underwriting an elimination rule for the concept of a thinking being:

TB-E From: x is a thinking being, infer: x is an LOT being.

We can also allow that it is part of our current conception that we ourselves are thinking beings. This underwrites an introduction rule for the concept of a thinking being:

TB-I From: x is one of us, infer: x is a thinking being.

For at least two reasons, appeal to this pair of an introduction rule and an elimination rule could not, by itself, provide a complete account of a fully determinate concept of a thinking being. One reason is that it is not plausible that the elimination rule, *TB-E*, should figure in an account of grasp of the concept of a thinking being. It is surely possible for someone to grasp that concept without embracing, or even considering, the philosophical theories that support Aunty's argument. A second reason is that these rules, together with the way the world is, do not determine an extension for the concept of a thinking being. They do not determine the status of an LOT being that is not one of us. However, my present concern is not with underdetermination of extension by rules, but with overdetermination.

Whether these inference rules are primitive or derived, if they are supposed to be truth preserving then, for certain ways the world might be, they do not permit any set to be assigned to the concept as its extension. In particular, if there is an object, x , that is one of us but not an LOT being, then *TB-I* requires that x is, while *TB-E* requires that x is not, a member of the extension of the concept of a thinking being. So, if the rules accurately reflect our current conception of a thinking being, and if our world is, in fact, a disobliging world in which we are not LOT beings, then our current conception does not correspond to a genuine concept.¹⁰

On this account, if the world is disobliging then there is strictly speaking a gap where the concept of a thinking being was supposed to be. If we were to discover that ours is a disobliging world, then we should negotiate our way towards a concept to fill that gap by revising our conception of what it is to be a thinking being. This process of revision would be informed by the particular ways in which the world had turned out to be disobliging. The revision of our conception would proceed under two constraints. The revised conception should be one under which we fall; so it should not involve a commitment to the truth of the LOT hypothesis.¹¹ And the revised conception should rationally sustain as much as possible of our folk psychological practice.¹² It is by acknowledging this pair of constraints on the process of revision that we honour the intuition of non-negotiability concerning our commitment to everyday folk psychological description, explanation, and prediction.¹³

2.3 Aunty's argument and the problem of armchair knowledge

If what I have just said is right then we can deflect the worry about eliminativism. But there is still a serious problem to be faced – indeed, it is the problem with which these lectures are primarily concerned. It may seem, then, that worrying about eliminativism has simply delayed the main event. But, when we come to solve the serious problem that we now face, the details of our response to the worry about eliminativism will be important.

Suppose that the philosophical theories supporting Aunty's argument really do provide the best clarification, elucidation and elaboration of our conception of a thinking being. Suppose that the conception really does underwrite the exemplar-based introduction rule, *TB-I*, and the more theoretically based elimination rule, *TB-E*. And suppose too that ours is an obliging world, so that the conception corresponds to a genuine concept and is not subject to revisionary pressure. Then it seems that philosophy's armchair methodology can yield knowledge that *I am a thinking being* and knowledge that *if I am a thinking being then I am an LOT being*. But, without rising from the armchair, I can surely put these two pieces of armchair knowledge together and perform a simple *modus ponens* inference to reach the conclusion that I am an LOT being:

- LOT(1) I am a thinking being.
- LOT(2) If I am a thinking being then I am an LOT being.
- Therefore:
- LOT(3) I am an LOT being.

The (LOT) argument is palpably valid but the epistemological commentary on it is problematic. We can agree that, if I have armchair knowledge that the premises LOT(1) and LOT(2) are true then I am committed, in the armchair, to believing that the conclusion LOT(3) is true, that is, to believing that I am an LOT being. But it is intuitively implausible that I could, from the armchair, come to know whether or not the

LOT hypothesis is true. It is implausible that I could, from the armchair, settle the question whether or not the LOT hypothesis is true – even whether it is true of me. Settling that question seems to require the empirical investigative methodology of cognitive science rather than the armchair methodology of philosophy.

Someone might suggest that the problematic result could be avoided if we were to deny that our conception of a thinking being underwrites the exemplar-based introduction rule. But even if this route to armchair knowledge that I am a thinking being were to be unavailable, I could still know it with first-person authority. Since at least some thinking is conscious, first-personal awareness of my own conscious mental states assures me that I am a thinking being. There is much to be said about the special character of our knowledge of our own conscious mental states. One thing that can be said is that the status of my special first-personal knowledge as knowledge does not depend on my having conducted any detailed empirical investigation of the world within me or around me (McLaughlin and Tye, 1998, p. 286). So, first-personal knowledge is armchair knowledge; and that is enough to generate the problem.

There is thus nothing to be gained – and, in terms of the worry about eliminativism, much to be lost – by denying that our conception of a thinking being underwrites the exemplar-based introduction rule. But we must acknowledge that there are two slightly different ways in which the problem of armchair knowledge can arise, depending on the exact nature of the warrant for believing LOT(1).

2.4 Philosophical methodology and a priority

Philosophy's methodology is an armchair methodology. But it would not be right to say that philosophy has a purely *a priori* methodology. One reason is that, even in the armchair, we presume upon commonplace assumptions that are empirical in character. A second reason – and the one that concerns me here – is that conceptual negotiation in a disobliging world has to take account of the ways in which the world has turned out to be disobliging. In the light of this second reason, we might regard philosophical theorising as involving two phases. The first phase is relatively purely *a priori*, though it may incorporate commonplace empirical assumptions. The second, more revisionary, phase is triggered by the discovery that some current conception is of no use to us because it does not correspond to any genuine concept, and this phase is characterised by greater involvement with the results of empirical research.

This picture of two separate phases is, no doubt, oversimplified. Philosophers who are ostensibly engaged in first-phase theorising – clarifying, elucidating and elaborating our current conception – are likely to give some consideration to the revisionary pressures that would result from one or another empirical discovery. They might anticipate directions that conceptual negotiation could take and they might try to formulate conceptions that could live at low risk, or even no risk, from discoveries in empirical science. But by distinguishing between the two phases we reduce the temptation simply to assume that our current conception is a low-risk, or no-risk, conception.

For example, it might be that a neo-behaviourist conception of thought and reasoning would be largely immune against discoveries in information-processing psychology or neuroscience. But that immunity against revisionary pressures does not show that philosophical theories of thought and reasoning that depart from behaviourism, like those that support Aunty's argument, are incorrect as accounts of the concepts that we actually use. Our current conception may be a high-risk conception.

3. An analogy: Colour concepts

Not everyone will find attractive the view of conceptions as having exemplar components and more theoretical components, and as living at risk. But I draw some encouragement from what David Lewis (1997) says about colour concepts.

If our concepts of colours and of colour experiences are concepts of properties of objects and of inner states that are simultaneously implicitly defined by our folk theory, then a philosophical analysis is likely to include such putatively definitional principles as these (1997, p. 327):

D1 *Red* is the surface property of things which typically causes experience of red in people who have such things before their eyes.

D2 *Experience of red* is the inner state of people which is the typical effect of having red things before the eyes.

The problem with D1 and D2 is that what they say, while true, does not distinguish the pair <red, experience of red> from other similar pairs such as <green, experience of green>. Something must be added to the folk theory of colour in order to individuate specific colours, such as red. Lewis suggests that there may be different versions of this additional component for different groups of users of the concept, each version specifying relatively parochial examples. Thus, amongst followers of Australian Rules football, it will suffice to say 'that red is the colour of the diagonal stripe on an Essendon Football Club jumper' (ibid., p. 335).

The theoretical component of our conception of red underwrites an elimination rule for the concept of red:

Red-E From: x is red, infer: there is a type of inner state that is typically caused in people who have x before their eyes.

The (parochial) exemplar component of my conception of red, underwrites an introduction rule for the concept of red:

Red-I From: x is the diagonal stripe on an Essendon jumper, infer: x is red.

But, for certain ways the world might be, these rules do not permit any set to be assigned to the concept of red as its extension. If our world is a disoblighing world in which there is not a single type of inner state that is typically produced in people by the Essendon stripe, then my (partly parochial) conception of red does not correspond to a genuine concept. In that case, we should negotiate our way towards a new colour concept to fill that gap. Presumably, this new concept would have to involve a measure of relativisation to individuals or to groups.

In this case, it seems clear that the option of being open to the possibility of conceptual revision in the light of empirical discoveries is preferable to two more extreme options. On the one hand, there is the option of standing ready to abandon colour talk altogether if the inner states caused by the Essendon stripe turn out to vary from person to person. On the other hand, there is the option of rejecting any philosophical theory about colours and colour experiences that includes principles such as D1 and D2. My claim is that the situation is similar in the case of the concept of a thinking being. The option of being open to the possibility of conceptual revision in the light of empirical discoveries is preferable to two more extreme options: on the one hand, standing ready to perform an eliminativist *modus tollens* and, on the other hand, rejecting any philosophical theory that imposes substantive empirical requirements for thought.

3.1 Colour concepts and the problem of armchair knowledge

Suppose that Lewis's account provides the best clarification, elucidation and elaboration of our conception of red, that this conception underwrites the introduction rule, *Red-I*, and the elimination rule, *Red-E*, and that ours is an obliging world. Then it seems that philosophy's armchair methodology can yield knowledge that *the Essendon stripe is red* and knowledge that *if the Essendon stripe is red then there is a type of inner state that is typically caused in people who have it before their eyes*. But then I can put these two pieces of armchair knowledge together to arrive at armchair knowledge that there is a single type of inner state that is typically caused in people who have the Essendon stripe before their eyes:

RED(1) The Essendon stripe is red.

RED(2) If the Essendon stripe is red then there is a type of inner state that is typically caused in people who have it before their eyes.

Therefore:

RED(3) There is a type of inner state that is typically caused in people who have the Essendon stripe before their eyes.

The (RED) argument, like the (LOT) argument, is clearly valid; but once again the epistemological commentary is problematic. For it is intuitively implausible that, just by deploying the armchair methodology of philosophy, I should be able to settle the question whether or not the Essendon stripe produces the same type of inner state in different people.

Someone might suggest that the problematic result could be avoided if we were to deny that my conception of red underwrites an exemplar-based introduction rule. But even if that denial were right, I could surely know that the Essendon stripe is red just by looking at an Essendon football jumper. There is much to be said about our perceptual knowledge of the colours of things. One thing that can be said is that the status of this perceptual knowledge as knowledge does not depend on my having conducted any detailed empirical investigation of people's inner states. So the problem remains. It is surely implausible that, without rising from the armchair save to look at an Essendon football jumper, I could settle the question whether or not the Essendon stripe typically causes the same type of inner state in different people.

Once again, there is nothing to be gained by denying that my conception of red underwrites an exemplar-based introduction rule. And, since the principles D1 and D2 do not distinguish the pair <red, experience of red> from <green, experience of green>, there is much to be lost. But we must acknowledge that there are two slightly different ways in which the problem of armchair knowledge can arise, depending on the exact nature of the warrant for believing RED(1).

4. The problem of armchair knowledge and the prospects for its solution

Aunty's argument gives rise to an instance of the problem of armchair knowledge. It also leads to a worry about eliminativism. In order to deflect the worry, I postulated a particular structure within our conception of a thinking being. The example of colour concepts was introduced to provide an analogy. Colour concepts, like the concept of a thinking being, answer to conceptions that include an exemplar component and a more theoretical component. So it is no surprise that colour concepts also give rise to an instance of the problem of armchair knowledge.

4.1 Externalism and armchair knowledge

The problem of armchair knowledge arises once again when we consider philosophical arguments to the effect that having thoughts with certain contents requires being embedded in a worldly environment in the right way – externalist arguments. While the case of architecturalist arguments is relatively unfamiliar, the case of externalist arguments is much better known.

Suppose that externalist philosophical theorising supports a conditional: If x is thinking that water is wet then x meets environmental condition E . Then the armchair methodology of philosophy can yield knowledge that *if I am thinking that water is wet then I meet environmental condition E*. Suppose that I am consciously thinking that water is wet. Then I can know, in the special first-personal way, that *I am thinking that water is wet*; and this too is armchair knowledge. I can put these two pieces of armchair knowledge together and perform a simple *modus ponens* inference to reach the conclusion that I meet environmental condition E :

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

WATER(2) If I am thinking that water is wet then I meet environmental condition E .

Therefore:

WATER(3) I meet environmental condition E .

But, for many choices of E , it is deeply implausible that I could settle the question whether or not I meet environmental condition E from the armchair.¹⁴

Just to dramatise the situation, we may suppose that externalist philosophical theorising supports the conditional: If x is thinking that water is wet then x is, or has been, embedded in an environment that contains samples of water. The problem then is that it is implausible that I could, from the philosopher's armchair, settle the question whether or not I am, or have been, embedded in an environment that contains samples of water.

4.2 Possible solutions to the problem of armchair knowledge

Both architecturalist arguments and externalist arguments give rise to instances of the problem of armchair knowledge. Some philosophers will draw a moral from this epistemological problem. Armchair philosophical theorising cannot, after all, take us from everyday folk-psychological claims about our thoughts and their contents either *down*, to substantive claims about the cognitive machinery that underpins our thinking, or *out*, to substantive claims about the world that our thoughts concern.

I cannot prove that this response is wrong, and it is certainly a tempting response if one has the impression that the problem arises only from a couple of idiosyncratic and easily rejected philosophical arguments. I have tried to avoid giving that impression. With the example of colour concepts, I have begun to show that the problem is rather more widespread than might at first have been suspected. I could go further in that direction by considering other architecturalist arguments and other arguments that uncover requirements for thinking thoughts with particular kinds of contents (such as indexical and demonstrative thoughts, or recognition-based thoughts).

In the case of externalist arguments, and possibly also in the case of Aunty's argument, the problem of armchair knowledge arises in part because of claims about our knowledge of our own conscious mental states. So a partial solution might be to contest first-person authority or privileged access. But this strategy does not generalise to all instances of the problem and, in any case, I am prepared to accept as a constraint on any

adequate solution to the problem that it should allow for specially authoritative self-knowledge.

Another possible solution would be to maintain that the problematic arguments all involve equivocation (Raffman, 1998). In the case of the (LOT) argument, the equivocation claim would be that, in one sense of ‘thinking being’, I can know from the armchair that I am a thinking being, but it is only in another sense of ‘thinking being’ that I can know from the armchair that if I am a thinking being then I am an LOT being. However, this putative solution is unmotivated given that the aim of philosophical theorising is to clarify, elucidate and elaborate the concepts that we actually use. The philosophical theories that support Aunty’s argument are put forward as accounts of the very concept of thinking being that I deploy when I think or say that I am a thinking being.

Instances of the problem of armchair knowledge turn on the implausibility of consequences to the effect that a question could be settled from the armchair – whether or not I am an LOT being, whether or not the Essendon stripe typically causes the same type of inner state in different people, whether or not I am embedded in an environment that contains samples of water. So, for a general solution, I must either deny that these consequences are really so implausible or else deny that they are really consequences.

The first, bullet-biting, option has its defenders, at least for the case of externalist arguments (Sawyer, 1998). Indeed, some philosophers aim to show that externalist arguments have anti-sceptical potential (Warfield, 1998). If armchair reflection allows me to settle the question whether or not there is water in my environment, perhaps it would also allow me to settle the question whether or not there is an external world at all – whether, instead, I am the envatted victim of a powerful but deceptive scientist. But as a general strategy this option does not appeal.¹⁵ However attractive it would be to have a response to the sceptic, it surely is implausible that we should be able to settle questions about our cognitive architecture without empirically investigating the information-processing machinery in our heads. So I shall be taking the second option. I deny that the implausible claims about settling questions from the armchair are really consequences.

Roughly, the strategy is this. I have to say that, from the facts that I can have armchair knowledge of a conditional and armchair knowledge of the antecedent of the conditional, it does not always follow that I can have armchair knowledge of the consequent of the conditional. So the solution to the problem of armchair knowledge lies in limitations on knowledge by inference. However, many complex issues surround the notion of knowledge. It may be, for example, that I can sometimes know *that* P even though I cannot do what we would ordinarily call settling the question *whether* P. In order to avoid some of these issues for the time being, I proceed in terms of epistemic warrant or justification rather than in terms of knowledge. I have to say that, sometimes, the epistemic warrants or justifications that I have for believing the premises of a palpably valid argument are not transmitted to the conclusion in a way that would allow me to settle the question whether or not the conclusion is true. So, even though I have armchair warrants for believing the premises of a palpably valid argument, it does not follow that I could, from the armchair, settle the question whether or not the conclusion is true.

There are many questions to be answered about this strategy for solving the problem of armchair knowledge. We shall come to some of them. But in the remainder of this first lecture I shall say something about the original source of the idea that there are cases of non-transmission of epistemic warrant.

5. Non-transmission of epistemic warrant: Wright on Moore

In 'Facts and certainty' (1985), Crispin Wright invites us to '[reflect] on the intuitive inadequacy of G.E. Moore's [1959] "proof" of the existence of the external world' (1985, p. 434). Moore's argument can be set out as follows:

MOORE(0) I am having an experience as of one hand [here] and another [here].

MOORE(1) I have hands.

MOORE(2) If I have hands then an external world exists.

Therefore:

MOORE(3) An external world exists.

MOORE(0) describes Moore's experience and this gives the impression that MOORE(1) is arrived at by inference. In the end, we shall want to allow that it is the experience itself, rather than a belief about the experience, that provides the defeasible warrant for believing the proposition about hands. But in any case, the question is whether this support for MOORE(1) is transmitted to MOORE(3) across the *modus ponens* inference in which the conditional premise, MOORE(2), is supported by an elementary piece of philosophical theorising.

Wright (1985, pp. 435–6) asks us to compare Moore's argument with the following:

ELECTION(0) Jones has just written an 'X' on that piece of paper.

ELECTION (1) Jones has just voted.

ELECTION (2) If Jones has just voted then an election is taking place.

Therefore:

ELECTION (3) An election is taking place.

Here, the evidence summarised in ELECTION(0) provides defeasible support for ELECTION(1); and this premise, together with ELECTION(2), which is warranted by the conceptual connection between voting and elections, clearly entails ELECTION(3). Given this relationship between ELECTION(1) and ELECTION(3), we might expect that empirical evidence against ELECTION(3) would count against ELECTION(1) by going into the scales on the opposite side from the evidence summarised in ELECTION(0). But, Wright stresses, this is not the correct picture (*ibid.*, p. 436):

Imagine . . . that you live in a society which holds electoral 'drills' as often as we hold fire drills, so that the scene you witness of itself provides no clue whether a genuine election is going on or not. In that case, unless you have further information, the knowledge that Jones has placed an 'X' on what looks like a ballot paper has no tendency whatever to support the claim that he has just voted.

In a situation where it is as likely as not that what I am watching is a drill rather than an election, the support ordinarily provided for ELECTION(1) by ELECTION(0) is not outweighed but removed. So, Wright says (p. 436):

the evidential support afforded by [ELECTION(0)] for [ELECTION(1)] is itself conditional on the *prior* reasonableness of accepting [ELECTION(3)]. . . . [K]nowledge of the first does not begin to provide support for the second unless it is *antecedently* reasonable to accept the third.

An imagined sceptic then says that Moore's argument is relevantly similar (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.

As a result (*ibid.*; emphasis added): ‘Only if Moore *already* has grounds for [MOORE(3)] does [MOORE(0)] tend to support [MOORE(1)].’

The sceptic’s point is that MOORE(3) cannot be supported indirectly by inference from particular claims like MOORE(1) any more than ELECTION(3) can be supported by inference from ELECTION(1). Direct, independent and antecedent support is what is needed. But while independent evidence in support of ELECTION(3) might be gathered, there is no prospect of such support for MOORE(3). If this pattern of sceptical argument is accepted then (*ibid.*, p. 438):

we seem bound to recognize that all our evidential commerce is founded upon assumptions for which we have no reason whatever, can get no reason whatever, and which may yet involve the very grossest misrepresentation of reality.

5.1 *Limiting transmission of warrant*

It is in the context of this discussion of a pattern of sceptical argument that Wright’s idea of non-transmission of epistemic warrant – in particular, non-transmission of evidential support – emerges (*ibid.*, p. 436–7):

It simply is not true that whenever evidence supports a hypothesis, it will also support each proposition which follows from it. The important class of exceptions illustrated are cases where the support offered to the hypothesis is conditional upon its being independently reasonable to accept one in particular of its consequences.

This clearly suggests a first shot at a principle limiting transmission of epistemic warrant – something along the following lines:

Limitation Principle (first shot)

Epistemic warrant is not transmitted from the premises of a valid argument to its conclusion if, for one of the premises, the support offered by the putative warranting element is conditional on its being antecedently and independently reasonable to accept the conclusion.

Here the neutral term ‘warranting *element*’ is intended to include within the scope of the principle both evidential warrants and perceptual warrants. The warrant for believing the premise of an argument might be provided by some further warranted belief or it might be provided by a perceptual experience.

No doubt, this principle stands in need of revision. But if Wright’s core idea were to contribute to a satisfying account of Moore’s argument then this would at least count against a blanket rejection of any limitation on the transmission of epistemic warrant. We could then try to use some revised and extended limitation principle to provide a solution to the problem of armchair knowledge.

5.2 *But is Moore’s argument an example of transmission failure?*

There is an important line of thought leading to the view that Wright’s core idea *does not* yield a satisfying account of Moore’s argument.

All the talk about acceptance of the conclusion of the argument needing to be antecedently and independently reasonable suggests that, in a fully explicit setting out of Moore’s argument, the conclusion MOORE(3) would appear as an additional premise or assumption, contributing to the derivation of MOORE(1). That, in turn, leads to the idea that the problem with Moore’s argument is that it is *implicitly circular*. If the argument is

implicitly circular, then it is not surprising that it should be intuitively inadequate for anti-sceptical purposes. Furthermore, on this construal, the argument invites a powerful sceptical response. The full story about my justification for believing that I have hands must include an antecedent and independent justification for believing that there is an external world – for believing that I am not, for example, the envatted victim of a powerful but deceptive scientist. But no such antecedent and independent justification for denying the sceptical hypothesis seems to be available.¹⁶

However, James Pryor (2000, forthcoming) argues that, although the sceptic can mount a powerful argument by relying on the implicit circularity claim, this claim is not true. Rather, perceptual experiences themselves provide defeasible justification for beliefs, such as the belief that I have hands, and this justification is immediate – it does not rest on any antecedent justification for any other belief (Pryor, 2000, p. 532).¹⁷ If Pryor is right about this and the warrant for MOORE(1) does *not*, after all, depend on there being an antecedent warrant for believing MOORE(3), then the limitation principle is not triggered. So Moore’s argument cannot serve as an example of non-transmission of warrant.

Indeed, the prospects for the idea of non-transmission of warrant seem to be even worse. For suppose that the sceptic were right and that MOORE(3) did need to figure as an additional premise or assumption, contributing to the derivation of MOORE(1). By the sceptic’s lights, no antecedent and independent warrant for that assumption is available, and so there would be no warrant for believing MOORE(1) on the basis of perceptual experience. This is the sceptic’s main point. So, on the sceptic’s account, Moore’s argument is still not an example of transmission failure, since there is no warrant to be transmitted in the first place.

We seem to have reached the conclusion that, whether the sceptic is wrong or right, Moore’s argument is not a case of non-transmission of epistemic warrant. So appeal to Moore’s argument cannot help me resist a blanket rejection of my strategy for solving the problem of armchair knowledge.

But I draw a rather different lesson from Wright’s seminal contribution. Although I accept most or all of what Pryor says about the justification of perceptual beliefs, I still agree with Wright that Moore’s argument is an example of non-transmission of warrant. It is apparent common ground between Pryor and Wright that there is a close relationship between non-transmission of warrant and begging the question. Indeed, Pryor says: (forthcoming, ms p. 3): ‘This notion of transmission-failure is basically a new piece of terminology for talking about an old phenomenon: the phenomenon of *begging the question*.’ So Pryor says that Moore’s argument does not beg the question, while Wright says that it does.

On this issue, I speak with Wright. But there is more than one notion of begging the question in the vicinity. So, the plausibility of my strategy for solving the problem of armchair knowledge depends on my making clear just what property of Moore’s argument – and of the (LOT), (RED), and (WATER) arguments – I am focusing on. Begging the question, and its relationship to transmission of warrant and to the notion of settling the question that figures in the problem of armchair knowledge, will be the topics for the second lecture.

Notes

¹ David Lewis notes that there is an element of tension in the methodology of philosophy (1986, p. 134): ‘I am trying to *improve* that theory [our total theory], that is to change it. But I am trying to improve *that* theory, that is to leave it recognisably the same theory we had before.’ There is always a judgement to be made as to whether clarification, elucidation and elaboration have taken us away from the concepts that we actually use.

² Lewis, 1989/2000, p. 85, says that a piece of philosophical analysis is supposed to be analytic, but might not be obviously analytic. He continues: ‘It is a philosophical problem how there can ever be unobvious analyticity. We need not solve that problem; suffice it to say that it is everyone’s problem, and it is not to be solved by denying the phenomenon. There are perfectly clear examples of it: the epsilon-delta analysis of an instantaneous rate of change, for one.’

³ Fodor says (1987, p. 139): ‘It turns out that dear Aunty is, of all things, a New Connectionist Groupie.’

⁴ Ramsey, Stich and Garon, 1990, seek to uncover a commitment to propositional modularity in our everyday explanatory practices. If they are right, then propositional modularity is embedded in our commonsense folk psychological theory. But they make no claim to be engaged in conceptual analysis, however broadly conceived.

Indeed, Stich, 1996, pp. 60–3, is deeply sceptical about the prospects for philosophical arguments that attempt to uncover ‘conceptually necessary’ requirements for being a thinking being. Stich complains that in my papers (Davies, 1990, 1991a, 1992a) ‘there is . . . not a single reference to Quine’ (1996, p. 88, n. 54). He says (ibid.): ‘In reading the books and articles of those who invoke the notion of a constitutive property, it is easy to get the feeling that one has fallen into a time warp. These philosophers write as though the notion of constitutive properties were entirely unproblematic, and they give no indication that they have ever heard of Quine and his assault on the analytic/synthetic distinction.’ But one could accept Quine’s criticisms of analyticity as truth purely in virtue of meaning without abandoning the idea that philosophy has a distinctive methodology that may reveal *a priori* requirements for being a thinking being.

For an argument that our explanatory practices do not require the assumption of propositional modularity, see Jackson, 1997.

⁵ In the original paper, I developed this step of the argument in two slightly different ways, drawing in turn on Evans, 1982, and on Peacocke, 1990. Here I follow the latter version. Peacocke’s notion of concept possession is presented in Peacocke, 1992, chapter 1; see p. 6 for the claim that certain inferences should be found primitively compelling in virtue of their form. In that book, Peacocke proposes that concepts can actually be individuated by their possession conditions, where those conditions are specified in terms of something like a functional (inferential) role. But this is not essential to Aunty’s argument.

⁶ See Davies, 1987, 1995. The account owes much to Evans, 1981.

⁷ Suppose that having a concept *C* involves commitment to a distinctive pattern of inference in which *C* occurs in the premises. Suppose, in particular, that the rule of inference is an elimination rule: From $\Sigma(C)$, infer Δ (where Δ depends systematically on Σ). Suppose that a transition-mediating mechanism, *M*, embodies tacit knowledge of this rule. Then *M* takes as inputs physical configurations underpinning thoughts whose contents have the form $\Sigma(C)$, and produces as outputs physical configurations

underpinning thoughts whose contents have the related form Δ , in which C does not occur. Since the operation of M provides a common causal explanation of the transitions that conform to the rule of inference, the physical configurations that underpin thoughts of the form $\Sigma(C)$ must have in common a physical property that encodes the occurrence of the concept C in the thoughts' contents.

⁸ Aunty's armchair argument involves a downward inference from the personal level of description to a subpersonal level of description. My commitment to architecturalist arguments, such as Aunty's, should be set against a methodological background provided by general views about inter-level relationships and inter-disciplinary relationships (Davies, 2000a, 2000b; see also Peacocke, 1992, chapter 7, esp. pp. 177–90).

On the relationship between the personal level of description and subpersonal levels, I favour the 'interaction without reduction' view. On the one hand, there are downward inferences. Philosophical theorising itself reveals that personal-level conditions impose subpersonal-level requirements. (This is what I mean by inter-level 'interaction'.) But, on the other hand, for some personal-level phenomena we find upward explanatory gaps when we try to construct illuminating accounts using just subpersonal-level notions. (This is why it is 'without reduction'.) To the extent that I expect upward explanatory gaps, my view differs from that of analytical functionalism.

If the personal-level concepts of consciousness, thought, action and meaning are susceptible of *a priori* functionalist analysis then facts about the instantiation and distribution of these mental phenomena will follow *a priori* from facts about the instantiation and distribution of physical properties. It does not follow from this that facts about these mental phenomena will also follow *a priori* from facts about information processing and cognitive machinery. So it is strictly possible that an analytical functionalist could agree with me that there are upward explanatory gaps between the subpersonal level of cognitive scientific description and the personal level of everyday mental description. But still, the functionalist is likely to expect upward *a priori* entailments from cognitive scientific facts to mental facts. Furthermore, if the central example of an upward explanatory gap is provided by the case of phenomenal consciousness, then someone who thinks that there is no prospect of a satisfying explanation in information-processing terms will also think that there is no prospect of a satisfying explanation in physical terms. Here, the contrast with functionalism is explicit.

On the inter-disciplinary relationship between, for example, philosophy of mind and language and the cognitive sciences, I favour a position that is intermediate between two extremes. I reject *philosophical isolationism*, which says that cognitive science has little or nothing to contribute to the philosopher's project of plotting the contours of our conceptual scheme. But I also reject *cognitive scientism*, which says that the only questions about mind and language that are susceptible of rational investigation are questions that belong to the sciences of mind and language. Philosophy has a distinctive methodology, different from that of the empirical sciences. But it still involves theory, and not just therapy.

⁹ Someone might say that the conclusion of the *modus tollens* inference would be only that we are not thinking beings as the neo-Fregean philosopher conceives of thinking beings. But, if the philosophical theories that support Aunty's argument provide the best clarification, elucidation and elaboration of our current conception of a thinking being, then the conclusion of the *modus tollens* inference is that we do not fall under the concept of thinking being that we actually use.

Ramsey, Stich and Garon say (1990, p. 500): ‘If connectionist hypotheses [of a particular sort] turn out to be right, so too will eliminativism about propositional attitudes.’ But the eliminativist argument as they envisage it proceeds from the premise that folk psychology is a theory that posits propositional attitude states. If certain kinds of connectionist hypotheses are true then this theory is false, and false in a way that would require theory change of an ‘ontologically radical’, rather than ‘ontologically conservative’ kind. So it would turn out that there are no such things as propositional attitudes. In the case of Aunty’s argument, the eliminativist *modus tollens* would proceed in a more straightforward way. If I am not an LOT being then I am not a thinking being. If we are not LOT beings then we are not thinking beings. If there are no LOT beings then there are no thinking beings.

¹⁰ There is an inevitable element of stipulation in my use of the contrast between ‘concept’ and ‘conception’ here. Having a conception of something is not just having a battery of views about that thing. It would be better to say that, normally, it is in virtue of having a conception that one grasps or possesses a concept. Stephen Schiffer, 2003, speaks of ‘underived conceptual roles’ that individuate concepts and of the possibility of there being a ‘tension’ in an underived conceptual role or a ‘glitch’ in a concept. I am not committed to the view that concepts are individuated by their underived conceptual roles. But much of what I say in terms of conceptions could be expressed in terms of conceptual roles.

¹¹ On the assumption that the philosophical theories supporting Aunty’s argument do provide the best clarification, elucidation and elaboration of our current conception, we need to revise that conception in order to avoid a commitment to the truth of the LOT hypothesis.

¹² We would not abandon the idea that we engage in deductive inference; but we would, presumably, adjust our conception of what is involved in accepting or performing an inference ‘in virtue of its form’.

¹³ For an analogy with the issues under discussion here, consider the views of Wiggins and Strawson on free will. Wiggins, 1973 (cf. van Inwagen 1975), offers a classical incompatibilist argument. His claim is that the best clarification, elucidation and elaboration of our current conception of free will reveals that free will requires that the thesis of determinism be false. If determinism is true then, by *modus tollens*, there is no free will and no one acts freely. Strawson, 1962, presses an intuition of non-negotiability. He argues that we cannot seriously envisage abandoning the ‘personal reactive attitudes’, whatever science might discover; and the practice of adopting personal reactive attitudes contains a slot for a notion of free action.

Strawson’s development of the intuition of non-negotiability does not, by itself, show that Wiggins has made an error in his argument for incompatibilism (Wiggins, 1973, p. 56): ‘How could a practical consideration – however all-embracing – prove the theoretical compatibility of freedom and determinism?’ Nor is it necessary for Wiggins to deny the intuition to which Strawson gives such vivid expression. Indeed, we can allow that our conception of freedom supports an incompatibilist argument, and that the intuition of non-negotiability goes beyond a practical consideration – that it is part of our conception that we are free agents. But if both those claims about our conception are correct, then there is no guarantee that any concept answers to our conception. (Schiffer, 2003, suggests that ‘our concept of free will has one component that inclines us to ascribe free will in certain conditions (“paradigm cases” of free acts, if there are any), whereas

another component disinclines us to ascribe free will when we know that the agent's decision was caused by factors over which he or she had no control'.)

Suppose that the world turns out to be a disobliging, because deterministic, world. Then we must negotiate our way towards a concept to fill the gap where the concept of freedom was supposed to be.

¹⁴ This instance of the problem of armchair knowledge is closely related to Michael McKinsey's argument for the claim that 'anti-individualism [externalism] is inconsistent with privileged access' (1991, p. 15).

¹⁵ I do not believe that externalist arguments, like those that support WATER(2), can provide us with a question-settling warrant for denying sceptical hypotheses. But I do not rule out the possibility that they may provide us with a reason for denying sceptical hypotheses. After all, my warrants for believing WATER(1) and WATER(2) do provide me with a reason to believe WATER(3). This suggests that there is a distinction to be drawn between ambitious and modest anti-sceptical projects, somewhat along the lines of Pryor, 2000, p. 517.

¹⁶ Wright, 1985, uses his reflections on Moore's argument in order to construct a sceptical argument (the 'second sceptical argument', see p. 434). This argument is very similar to the argument that Pryor develops on behalf of the sceptic (2000, section 2).

¹⁷ Perceptual experiences provide justifications for beliefs about hands and for beliefs about colours and shapes (properties that the perceptual system itself represents as such). It may be that the justificatory role of perceptual experience is different in these two kinds of case. But, even if the justificatory story for beliefs about hands is more complicated than the justificatory story for beliefs about shapes, this does nothing to re-introduce the idea of an antecedent warrant for believing that there is an external world.