

Aunty's argument and armchair knowledge

MARTIN DAVIES

In my contribution to the Proceedings of the First International Colloquium on Cognitive Science, held in Donostia (San Sebastián) in 1989, I advanced what I called 'Aunty's own argument for the language of thought'.¹ The Aunty in question is Jerry Fodor's. He represents her as a conservative figure who is more likely to favour connectionism than to accept that there are good reasons to adopt the language of thought hypothesis.² As I envisaged her, she has some sympathy for the views of the later Wittgenstein but is fundamentally a neo-Fregean. I claimed that the neo-Fregean framework offers Aunty the resources to construct her own argument for the language of thought hypothesis, an argument that is relatively non-empirical in character.³

Aunty's own argument is an argument for the claim that conceptualised thought requires the truth of the language of thought (LOT) hypothesis. It 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 patterns of inference. In particular, it is assumed that conceptualised thought involves performing certain inferences *in virtue of their form*. The second step makes use of a quite general connection between tacit knowledge of rules and syntactically structured representations. A background assumption for the whole argument is 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. 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 the assumption of *propositional modularity*.⁴ In my view, we are committed to this assumption by some of our everyday practices of mental talk and explanation, but I shall not spell out the nature of this commitment here.

In this paper, I briefly review Aunty's argument and then respond to an objection, address a worry, and propose a solution to a serious problem.⁵

1. A brief review of Aunty's argument

In order to think particular types of thought (such as the thought that $7+5=12$) a thinker must possess particular concepts (here, the concepts of addition and of equality and concepts of the numbers 5, 7 and 12). Furthermore, possessing particular concepts involves a thinker in commitments to particular forms of inference. This is not to say that there are forms of inference that the thinker will invariably accept when they are presented or perform when they are appropriate. Possessing a concept is a matter of having certain cognitive abilities, but there are many factors that may prevent an ability from being exercised. In Chomsky's (1965) terminology, possessing a concept is a matter of competence rather than of performance.

¹ Davies, 1992.

² Fodor, 1987, p. 139: 'It turns out that dear Aunty is, of all things, a New Connectionist Groupie.'

³ Cf. Rey, 1995.

⁴ See Ramsey, Stich and Garon, 1990.

⁵ The objection, the worry, and the problem arise even if we consider, instead of the language of thought hypothesis, the conditional statement: If intentional realism is true then the LOT hypothesis is true.

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.⁶ 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.⁷ What it is for a rule to be tacitly known, in the sense that figures in Aunty's argument, can be elucidated in terms of a structure in the explanations of causal processes.⁸ The causal processes to be considered are transitions between representations. Thus the inputs to, and outputs from, the processes are physical configurations that have semantic properties. For example, the input configurations might represent letter strings, and the output configurations might represent pronunciations. Given such a process, there may be a pattern in the input-output relation when the inputs and outputs are considered under their semantic (representational) descriptions. Thus, for example, it might be that whenever the input configuration represents a string beginning with the letter 'b' the output configuration represents a pronunciation beginning with the sound /B/. In such a case, we can say that the input-output transitions *conform* to a rule about the task domain; in the example, this would be the rule that letter strings beginning with 'b' have pronunciations beginning with /B/. But to say that the transitions conform to the rule is not yet to say that the mechanism that mediates those transitions embodies tacit knowledge of that rule. What is required for tacit knowledge of the 'b'-to-/B/ rule is that the transitions that conform to the rule should have a *common causal explanation*. This condition is met if there is, within the overall transition-mediating mechanism, a component processor or module that operates as a *causal common factor* to mediate all the transitions that instantiate the 'b'-to-/B/ pattern.

The first step in Aunty's argument says that thinking requires concept possession and concept possession requires tacit knowledge of rules of inference. The second step then makes use of the point that tacit knowledge of a rule requires syntactic articulation in the physical configurations that are the inputs to the transitions that are explained by the presence of that tacit knowledge.

⁶ 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.

⁷ Polanyi, 1967.

⁸ The account owes much to Evans, 1981. See Davies, 1987, for some details omitted here and Davies, 1995, for application of the notion to connectionist networks.

Consider again the device that takes representations of letter strings as inputs and produces representations of pronunciations as outputs. Although letter strings have internal structure, there is no requirement as yet that the input representations should have internal structure; something unstructured can represent something structured. But suppose that the device embodies tacit knowledge of the 'b'-to-/B/ rule. The 'b'-to-/B/ transitions all have a common causal explanation; there is a component processor that operates as a causal common factor to mediate all those transitions. In that case, the various input configurations that represent strings beginning with the letter 'b' need to share some property that will engage or activate the 'b'-to-/B/ component processor. This will be (i) a physical property that (ii) is correlated with the semantic property that these input representations share (that they all represent strings beginning with the letter 'b') and (iii) is a determinant of the input configuration's causal consequences. In short, this property will meet the minimal conditions for being a *syntactic* property.⁹

If the device also embodies tacit knowledge of the 'i'-to-/I/ rule and the 'n'-to-/N/ rule, then an input representation of the letter string 'bin' must have three syntactic properties. These will be correlated with the three semantic properties of representing a string beginning with the letter 'b', representing a string with 'i' in the middle, and representing a string ending in 'n'. In fact, quite generally and quite independently of any consideration of the LOT hypothesis, transition-mediating mechanisms that embody tacit knowledge of rules must have syntactically structured input representations.

We have not taken it as a background assumption that the physical configurations that underpin thoughts should have internal structure. It is consistent with intentional realism or propositional modularity that an occurrence of the thought that $7+5=12$ should be underpinned by activation at a single unit (the neural analogue of a light bulb being switched on). But, by the second step of Aunty's argument, any physical configuration that figures as input to component processors that embody tacit knowledge of rules of inference must be syntactically structured.

Thus, consider an occurrence of a thought in whose content the concept C is a constituent. Suppose that possession of the concept C requires commitment to at least one pattern of inference (where C occurs in the premises)¹⁰ so that, by the first step of Aunty's argument, the thinker meets the conditions for tacit knowledge of the corresponding rule of inference. Then, to the extent that the thought, as it occurs on this occasion, is apt to figure as a premise in an inference of this form, the physical configuration that underpins the thought must have a syntactic property that encodes the occurrence of the concept C in the thought's content. When several concepts are constituents in the content of a thought, and the thought is apt to figure as a premise in inferences of various forms associated with possession of those concepts, the physical configuration corresponding to the thought must have several syntactic properties. These will be physical properties that encode the occurrence of the various concepts as constituents in the thought's content and, indeed, encode the structural relations amongst those constituents. Thus, Aunty's argument takes us from the initial assumption of intentional realism to a version of the LOT hypothesis.

⁹ See Fodor, 1987, pp. 16–21.

¹⁰ The argument presented in the text would not apply if the only rules of inference associated with the concept C were introduction rules.

2. Response to an objection: ‘No need to descend’

Aunty’s argument involves a shift between levels, a *downward inference* from the personal to the subpersonal level. We start with personal-level notions of judgement, thought and inference and move to subpersonal-level notions of physical configuration and transition-mediating mechanism, notions that belong in the science of information-processing psychology. Someone who rejects the idea that our conception of ourselves as thinking beings involves a commitment to the truth of the LOT hypothesis may take this shift between levels as the target for an objection. In particular, it may be said that, where a philosophical theory supports such an inference from the personal to the subpersonal level, this is the product of an assumption that the inter-level relation is one of *reduction*. If persons are conceived in purely mechanistic terms then inevitably there will be subpersonal-level requirements (such as the truth of the LOT hypothesis) for personal-level conditions (such as conceptualised thought).

But, according to the imagined objector, putative downward inferences will be undermined once we take proper account of our distinctive nature. We are conscious beings, our thinking is subject to normative requirements, we arrive at judgements and decisions on the basis of reasons, and rationalising explanations of our beliefs and actions in terms of reasons are of a kind quite different from scientific explanations. This distinctive nature cannot be adequately described in the terms favoured by information-processing psychology (nor in the terms of neuroscience, biology, chemistry or physics). If we insist on a non-reductionist conception of ourselves as conscious thinking subjects and agents then we shall see that such requirements as there are on personal-level conditions can be met from the rich resources of the personal level itself. There is *no need to descend* to the subpersonal level of information-processing psychology.¹¹

So far, the envisaged objection is entirely schematic. One way of developing it is to argue, first, that a thinker needs to be consciously aware of the forms of the inferential transitions that he performs and, second, that this personal-level requirement undermines the argument for syntactic structure at the subpersonal level. If the form of the inference is already available to the thinker in conscious awareness then why, the objector asks, does it need to be encoded in physical configurations as well?

The objector insists that transitions in thought are subject to norms of rationality. What goes along with this insistence is the idea that if thought is to yield knowledge it must measure up to requirements of rational justification; reliable correlation between thought and reality is not enough. One fairly modest component in a not-purely-reliabilist epistemology is that knowledge requires some appreciation of how one could be getting things right.¹² When a belief is arrived at by an inferential transition that instantiates a valid form, there are some obvious constraints on how this idea is to be elaborated. The thinker cannot be required to use an overarching justificatory principle as an additional premise in his reasoning; such a requirement would be regressive. It would also be wrong to require that the thinker provide a theoretical account of the valid form that his inference instantiates; such a requirement would be too strict. But, it seems plausible that, in order to have an appreciation of how he could be getting things right, the thinker should, in some way, be aware of the form of his inference.

This requirement of *awareness of logical form* goes beyond what is needed for tacit knowledge of rules of inference, for tacit knowledge of rules may be present in a system

¹¹ This objection is considered at greater length in Davies, 2000a.

¹² See Brewer, 1995, 1996, 1999.

that lacks consciousness altogether. This is as it should be for a notion of tacit knowledge that is to serve the general purposes of information-processing psychology and cognitive science. But we can agree with the objector that, when we are concerned with knowledge and rationality in conscious thinking subjects, a total lack of awareness is not what is wanted. According to Aunty's argument, the presence of tacit knowledge of rules of inference and the truth of the LOT hypothesis are necessary conditions for conceptualised thought. There is no suggestion that they are sufficient conditions and we should now make it explicit that they are not.

The objector, not satisfied with this concession, maintains that, once the personal-level requirement of awareness of logical form is acknowledged, this actually undermines the downward inference in Aunty's argument. In response, I reject the suggestion that proper attention to our distinctive nature reveals that there is no need to descend to the subpersonal level. Acceptance of downward inferences is consistent with a non-reductionist conception of ourselves.

Aunty's argument begins from a premise about a thinker performing inferential transitions in virtue of their form. The notion of 'in virtue of its form' is, at least in part, causal: the form of an inference is to figure in the causal explanation of a thinker's performing the inferential transition in thought. The two steps of the argument first unpack this causal condition in terms of tacit knowledge and then make use of the general connection between tacit knowledge and syntactic structure. The objector has motivated the claim that the thinker should be aware of the logical form of the inference and I have accepted her claim. This is certainly a personal-level, rather than a subpersonal-level, requirement. But clearly, a thinker's being aware of the form of an inference that he performs is not yet sufficient for the form's figuring in the causal explanation of his performing that inference. So the personal-level requirement of awareness of form cannot undermine an argument that aims to show that there are subpersonal-level requirements for the personal-level condition of performing inferential transitions in virtue of their form.

The objector may say that I have not treated her objection fairly. She may say that it was no part of her position that performing an inferential transition and being aware of its form can supplant the notion of performing an inference in virtue of its form. Rather, her point was that awareness of form should be incorporated into a personal-level account of performing an inference in virtue of its form and that this personal-level account could then compete with the subpersonal-level account cast in terms of tacit knowledge of a rule of inference. And, she may say, in this competition the personal-level account would be at an advantage since causal transitions whose nature is utterly hidden from the thinker do not measure up to the requirements for rational, knowledge-yielding thought.

If this is the objector's point then there is some justice in it. But it is not correct to suppose that there can be a straightforward competition between a personal-level and a subpersonal-level account of the at-least-partly-causal notion of performing an inference in virtue of its form. After all, the key idea for the first step in Aunty's argument is that the best philosophical elucidation of performing an inference in virtue of its form will reveal that it involves meeting the conditions for tacit knowledge of the corresponding inferential rule. In response to this the objector may say that the conditions for tacit knowledge, cast in terms of transitions of the same form having a common causal explanation, can themselves be met at the personal level, so that there is still no need to descend to the subpersonal level of information-processing psychology. But what the objector needs to establish is that, even given the background assumption of intentional

realism or propositional modularity, the conditions for tacit knowledge could be met at the personal level and not met at the subpersonal level of information-processing psychology. We have been given no reason to suppose that this can be established.

3. The worry about eliminativism: An intuition of non-negotiability

Aunty's argument supports a conditional: If we are thinking beings then the LOT hypothesis is true of us; that is, we are LOT beings. Although the argument is relatively non-empirical in character, the question whether we really are LOT beings is a substantive empirical one and answering it requires detailed empirical investigation.¹³ Many empirical arguments have been advanced to support the LOT hypothesis but the issue is contested. A good deal of connectionist research aims to show that the kinds of behaviour that have been taken to constitute evidence in favour of the LOT hypothesis may be forthcoming from connectionist networks. Furthermore, it is argued that the networks in question do not meet the conditions for tacit knowledge of rules or syntactically structured representations either at the proprietary level of network description in terms of units and connections, activations and weights, or at any higher level of description. It seems reasonable to allow that it is epistemically possible (whether or not it is likely) that we may turn out not to be LOT beings. But then Aunty's argument would support an eliminativist *modus tollens*. From the premise that we are not LOT beings we would be able to conclude that we are not thinking beings.¹⁴

Imagine, for a moment, that empirical evidence decisively supported the thesis that we are not LOT beings. It seems that, in those circumstances, we would face a stark choice between two alternatives. On the one hand, we could perform the *modus tollens* inference and cease to regard each other and ourselves as thinking beings. On the other hand, we could conclude that there is something wrong with Aunty's argument. But the first alternative seems rationally to require that we abandon our familiar descriptions of ourselves and others as believing and wanting things, as hoping and fearing things, as engaging in reasoning and planning; and there are powerful intuitions proclaiming that this option is not genuinely available to us. Our everyday engagement in folk psychological practice seems to be philosophically non-negotiable. So, we would be driven to the second alternative. If we found ourselves to be in a disobliging world then we would be bound to reject Aunty's argument. We would have to conclude that the philosophical theories that support the argument are in some way flawed.

But, if this is the conclusion that we should draw if things turned out badly for the LOT hypothesis, then should we not draw that conclusion now? For the credentials of a philosophical and non-empirical argument ought not to depend on what our internal information-processing machinery turns out to be like. Given the intuition of non-negotiability concerning our engagement in folk psychological practice, it appears that the very possibility of an eliminativist *modus tollens* already counts against Aunty's argument. Indeed, that intuition of non-negotiability seems to count against any argument that purports to uncover substantive requirements for thought if there is a prospect that empirical investigation may reveal that those requirements are not met.

However, against all these considerations we must set the fact that a philosophical theory that avoids the threat of eliminativism by imposing no substantive empirical

¹³ This is so even given the assumption that intentional realism is true of us.

¹⁴ Cf. Ramsey, Stich and Garon, 1990, p. 500: 'If connectionist hypotheses [of a particular sort] turn out to be right, so too will eliminativism about propositional attitudes.'

requirements for thought is itself counterintuitive. The threat of eliminativism is posed by any theory that imposes requirements that concern internal information processing. The threat is avoided by theories that impose no requirements that go beyond readily observable patterns in behaviour. But theories that make thought supervene on behavioural trajectory face familiar counterexamples. We can imagine systems that produce the right kind of behaviour but are no more capable of thought than is a puppet or a jukebox.¹⁵

Standing ready to perform an eliminativist *modus tollens* goes against the non-negotiability of our engagement in folk psychological practice. Restricting our philosophical theories to those that totally avoid the threat of eliminativism has counterintuitive consequences. Neither option is attractive. But in my view we can respect the intuition of non-negotiability even while embracing the philosophical theories that support Aunty's argument.

We can accept that those philosophical theories provide the best elaboration and precisification of our current conception of a thinking being and that Aunty's argument correctly draws out a necessary condition for falling under that conception. But we can also allow that it is part of our current conception that we ourselves are thinking beings: being one of us is a sufficient condition for falling under that conception. Suppose that these claims about a necessary condition and a sufficient condition for falling under our current conception of a thinking being are both correct. It follows that if we are not LOT beings then our current conception dictates both that we are and that we are not thinking beings. In a disobliging world, our current conception of a thinking being would be of no use to us, since it would dictate contradictory answers to the question whether we are thinking beings.

If we turn out not to be LOT beings then we must negotiate our way to a revised conception of what it is to be a thinking being.¹⁶ This conceptual negotiation 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.¹⁷ And the revised conception should rationally sustain as much as possible of our folk psychological practice.¹⁸ It is by acknowledging this pair of constraints on the process of revision that we honour the intuition of non-negotiability concerning our engagement in folk psychological practice.

In response to the worry about eliminativism, what is being proposed is that the concept of a thinking being has at least two components. There is an exemplar component that specifies sufficient conditions: we, at least, are thinking beings. And there is a more theoretical component which, according to Aunty's argument, imposes a necessary condition: thinking beings are LOT beings. There is no logical guarantee that the items that meet the sufficient conditions also meet the necessary condition and in a disobliging world the two components lead to contradictory verdicts on cases.

¹⁵ See Peacocke, 1983; Block, 1981, 1995.

¹⁶ The process of revision will be informed by the particular ways in which the world turns out to be disobliging.

¹⁷ On the assumption that the philosophical theories supporting Aunty's argument do provide the best elaboration and precisification of our current conception, we need to revise that conception in order to avoid a commitment to the truth of the LOT hypothesis.

¹⁸ 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.

This proposal is somewhat analogous to what David Lewis (1997) says about colour concepts. He begins from the thought that our folk theory of colours contains principles linking colours and colour experiences, such as: ‘When a red thing is before someone’s eyes, it typically causes in him an experience of redness.’¹⁹ 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 theory about colours and colour experiences is likely to include such principles as these:

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.²⁰

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’.²¹

According to the (parochial) exemplar component of the concept of red, being the colour of the Essendon stripe is sufficient for being red: Essendon stripes (at least) are red things. From the theoretical component we can derive a necessary condition for being red: If something is red then there is a type of colour experience and a type of inner state that is typically caused in people who have that thing before their eyes. But there is no logical guarantee that there is a single type of colour experience and a single type of inner state that is typically produced in people by the diagonal stripe on an Essendon jumper. If there is not, then the two components of the concept yield contradictory pronouncements. If the world turns out to be disobliging in this respect then we must negotiate our way to revised, presumably relativised, colour concepts.

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 colour experiences and the corresponding inner states 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, similarly in the case of the concept of a thinking being, the option of being open to the possibility of conceptual revision is preferable to the more extreme options of standing ready to perform an eliminativist *modus tollens* or rejecting any philosophical theory that imposes substantive empirical requirements for thought.

4. A serious problem: Armchair knowledge

So far, I have responded to the ‘no need to descend’ objection and addressed the worry about eliminativism. But there is still a serious problem for Aunty’s argument.

¹⁹ Lewis, 1997, p. 326.

²⁰ Ibid., p. 327.

²¹ Ibid., p. 335.

Suppose that the LOT hypothesis is true and that the concept of a thinking being is in good order. It seems that, by relying on my grasp of the exemplar component of the concept of a thinking being, I can know that I am a thinking being. In fact, it seems that I have more than one way of knowing this. Since at least some thinking is conscious, first-personal awareness of my own conscious mental states also assures me that I am a thinking being. Either way, provided that the LOT hypothesis is in fact true, this knowledge seems to be available to me ahead of any empirical investigation of the information-processing mechanisms inside my head.

By relying on my grasp of the theoretical component of the concept of a thinking being, engaging in some inferences to the best philosophical explanation, and following through Aunty's argument I can, if the argument is a good one, come to know that a thinking being must be an LOT being. I know that if I am a thinking being then I am an LOT being.

Without conducting any detailed empirical investigation, I can have two pieces of knowledge that provide the premises for a simple *modus ponens* inference:

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.

But if, given knowledge of the premises, simply performing the inference yields knowledge of the conclusion then I can know that the LOT hypothesis is true (of me, at least) without any detailed empirical investigation. This seems counterintuitive. If the LOT hypothesis is indeed true then knowledge of that fact will be the result of experiments, computational modelling and, more generally, detailed comparison of the successes and failures of competing research programmes.

Aunty's argument supports a conditional, LOT(2), that gives rise to an instance of *the problem of armchair knowledge*. Both LOT(1) and LOT(2) can be known from the armchair, yet knowledge of LOT(3) surely requires an investigative methodology rather than an armchair methodology.

Another instance of the problem of armchair knowledge arises from my proposal (following Lewis) about colour concepts. Suppose that our colour concepts are in good order and consider the following *modus ponens* inference:

RED(1) This [pointing at the diagonal stripe on an Essendon jumper] is red.

RED(2) If this is red then there is a type of colour experience and a type of inner state that is typically caused in people who have this before their eyes.

Therefore:

RED(3) There is a type of colour experience and a type of inner state that is typically caused in people who have this before their eyes.

By relying on my mastery of the exemplar component of the concept of red (the parochial exemplar component that applies to my group), I can know that this Essendon stripe is red. Indeed, I have more than one way of knowing this since I can often know what colour something is just by looking at it. Having seen many Essendon jumpers, I can recognise this item as being the colour of the Essendon diagonal stripe. Either way, knowledge of RED(1) is available to me ahead of any investigation of other people's colour experiences or inner states. By relying on my grasp of the theoretical component of the concept of red (including the principles D1 and D2), I can know that if something

is red then there is a type of colour experience and a type of inner state that is typically caused in people who have that thing before their eyes. So I can know RED(2). But it is implausible that, without rising from the armchair save perhaps to look at an Essendon football jumper, I can know the conclusion RED(3).

The same kind of problem also arises from externalist theories about the intentional content of mental states. Sitting in my armchair, I can know that I am thinking that water is wet. Sitting in my armchair and following through some of the arguments in externalist theories, I can, if those theories are correct, know that if I am thinking that water is wet then I am, or have been, in an environment containing samples of water. Sitting in my armchair, I can draw the obvious conclusion. Yet it is counterintuitive that a mere exercise in armchair reflection should provide me with knowledge that I am, or have been, in a watery environment.²²

The problem of armchair knowledge is serious and it may be thought to cast doubt on the philosophical theories that give rise to it and, in particular, on the theories that support Aunty's argument for the language of thought. But in my view the problem is soluble. The key to the solution is that armchair epistemic warrants for the premises of an argument are not always transmitted to the conclusion of the argument, even when the inference is palpably valid. There are limitations on the transmission of epistemic warrant from premises to conclusion.²³

5. A solution: Limitations on transmission of epistemic warrant

My proposals for principles that limit the transmission of warrant are motivated by the claim 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. In order to make use of this claim, we need to have an account of begging the question and for this I rely on Frank Jackson (1987). He says that an argument begs 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'.²⁴

According to Jackson's view of what is achieved by advancing an argument for a conclusion, the speaker invites the hearer to borrow evidence, or other considerations, in favour of the premises of the argument. By her choice of premises she provides an indication as to what kind of considerations these are. Typically, evidence counts in favour of a proposition only relative to particular background assumptions and often the relevant background assumptions are shared between speaker and hearer. But when background assumptions are not shared, it is possible that the considerations that count in favour of the premises relative to the speaker's background assumptions do not count in favour of the premises relative to the hearer's background assumptions. Suppose that a speaker sets out to convince a doubting hearer of the truth of some conclusion. The speaker begs the question against the hearer if the hearer's doubt rationally requires him to adopt background assumptions relative to which the considerations that are supposed to support the speaker's premises no longer provide that support. A question-begging argument 'could be of no use in convincing doubters'.²⁵

²² See Davies, 2000b.

²³ For an earlier development of this idea, see Davies, 1998.

²⁴ Jackson, 1987, p. 111.

²⁵ *Ibid.*, p. 112.

Convincing a doubter and providing an epistemic warrant both involve ruling out various ways in which a proposition could have been false. A speaker's evidence for her premises rules out various ways in which those premises could have been false, ways left open by the speaker's background assumptions. A hearer who doubts the conclusion of the argument may have background assumptions that leave a wider range of possibilities open, and the speaker's evidence for the premises may not rule all those possibilities out. Indeed, the speaker's evidence may leave untouched the ways in which, according to the hearer, the conclusion could be false.

In a similar way, a thinker's epistemic warrants for believing the premises of an argument rule out various ways in which those premises could have been false, ways left open by background assumptions that the thinker is, in that context, epistemically entitled to make. But it does not follow that those same warrants rule out ways in which the conclusion could have been false which are left open by assumptions that the thinker is entitled to make. It may be, for example, that in the context of considering the conclusion the thinker is entitled to make fewer assumptions than when only the premises were under consideration. In that case, epistemically adequate warrants for the premises may be inadequate to rule out the possibilities that now need ruling out. On the other hand, it may be that the assumptions that the thinker is entitled to make already leave open no ways in which the conclusion could have been false. In that case, the thinker is epistemically entitled to believe the conclusion but the epistemic warrants for believing the premises contribute nothing to that entitlement. In either case, the thinker's epistemic warrants for believing the premises do not add up to an epistemically adequate warrant for the conclusion.

The analogy between convincing a doubter and providing an epistemic warrant motivates the following principle:

First Limitation Principle

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 order to apply this principle to any particular argument, we need to identify assumptions such that, for one of the premises, it is only against the background of those assumptions that the epistemic warrant for the premise counts as a warrant. Then we need to show that acceptance of those assumptions cannot be combined with doubt about the truth of the conclusion. There do seem to be cases that fit this pattern.²⁶ But it is not clear how we can apply the principle to explain the failure of transmission of warrant from LOT(1) and LOT(2) to LOT(3).

In the case of the conditional premise LOT(2), it might reasonably be said that it is only against the background of the assumption of intentional realism that the premise is supported by neo-Fregean philosophical theory. But it is surely not true that acceptance of this assumption cannot be rationally combined with doubt about the conclusion LOT(3). In the case of the first premise, LOT(1), the epistemic warrant is constituted either by grasp of the exemplar component of the concept of a thinking being or else by awareness of my own conscious mental states. But in neither case is there an obvious

²⁶ See the discussion of Moore's anti-sceptical argument in Davies, 1998, 2000b; and cf. Wright, 1985.

candidate for the role of background assumption without which the epistemic warrant would not count as a warrant.

There is, however, a very basic assumption that lies in the background of any epistemic project; namely, the assumption that there is such a proposition as the proposition for which one is attempting to provide evidence, justification or warrant. The notion of a proposition that figures in this assumption is not to be construed in a metaphysically committed way. If a thinker is attempting to provide a warrant for believing A then the basic background assumption is simply that there is such a thing to think as A. If there were no such thing to think as A then there could be no question of anything constituting an epistemically adequate warrant for believing A. So we can make explicit a second principle that is arguably a consequence of the first:

Second Limitation Principle

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.

One way in which the assumption that figures in this principle could turn out to be false would be if one of the purported conceptual constituents in the premise were revealed to be internally incoherent, dictating contradictory answers to the question whether some particular item falls under the concept. To that extent, the principle holds some promise of providing a solution to the instance of the problem of armchair knowledge that is posed by Aunty's argument. But, on the other hand, it is clear that acceptance of the assumption that there is such a thing to think as that I am a thinking being and, in particular, acceptance of the assumption that the concept of a thinking being is in good order can be rationally combined with doubt about the truth of the LOT hypothesis. It is only the acceptance of Aunty's argument that generates rational tension between acceptance of the background assumption and doubt about the conclusion.

It is clear what kind of modification of the principle is required if it is to provide a solution to the problem of armchair knowledge that arises from Aunty's argument and the modification is not merely opportunistic or *ad hoc*. To see this, we need to return to begging the question and focus on the fact that arguments may have several premises. Suppose that a speaker advances a multi-premise argument in an attempt to convince a hearer who doubts that argument's conclusion. The speaker offers various considerations for borrowing; they are considerations that count in favour of the premises relative to the speaker's background assumptions. If the hearer's doubt by itself rationally requires him to adopt background assumptions relative to which one of the speaker's premises is no longer supported by the considerations that she offers for borrowing then the speaker begs the question against the hearer. That is the kind of case that Jackson describes.

But there is a more complicated scenario in which it is no less true that the argument, as advanced by the speaker, will be of no use in convincing the doubting hearer. If the hearer is to be convinced then he must accept the considerations that the speaker offers in support of her premises. In addition, he must not differ from the speaker in his background assumptions in such a way that the premises are no longer supported by those considerations. Suppose that the hearer's doubt about the conclusion, when put together with acceptance of the considerations that the speaker offers in support of some of the premises, rationally requires him to adopt background assumptions relative to which some other one of the premises is no longer supported by the considerations offered in support of it. That is enough to ensure that the argument, as advanced by the speaker, will

be of no use in convincing the hearer. So, if failure of transmission of epistemic warrant is the analogue, within the thought of a single subject, of the dialectical phenomenon of begging the question then we should expect the following pair of limitation principles, of which the second is arguably a consequence of the first:

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.

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 is this last principle that provides a solution to the instance of the problem of armchair knowledge that is posed by Aunty's argument.

Suppose that a thinker accepts that there is such a thing to think as the premise LOT(1), that he himself is a thinking being. Suppose, in particular, that he accepts that there is no internal incoherence, no source of contradictions, in the concept of a thinking being. In that case, the thinker must accept the assumption that the items, such as himself, that meet the sufficient conditions for falling under the concept also meet the necessary conditions. Acceptance of that assumption does not, by itself, rationally preclude doubt about whether he himself is an LOT being. But suppose, in addition, that the thinker accepts the epistemic warrants for the premises LOT(1) and LOT(2). The epistemic warrant for believing LOT(1) is provided either by the exemplar component of the concept of a thinking being or else by his awareness of his own conscious mental states. But it is the warrant for believing the conditional premise LOT(2) that figures crucially in the solution to the problem of armchair knowledge. That warrant is provided by a battery of philosophical theory and by Aunty's argument. Acceptance of the assumption that the items that meet the sufficient conditions for falling under the concept of a thinking being also meet the necessary conditions *and of the warrant for LOT(2)* cannot be rationally combined with doubt about whether the thinker himself is an LOT being. So the limitation principle is triggered and epistemic warrant cannot be transmitted from LOT(1) and LOT(2) to the conclusion LOT(3). The warrants for believing the premises do not add up to an epistemically adequate warrant for the conclusion.

Because there are limitations on the transmission of epistemic warrant, armchair knowledge of the premises LOT(1) and LOT(2) does not lead to armchair knowledge of the conclusion LOT(3). Aunty's argument does not, after all, have the counterintuitive consequence that I can know that the LOT hypothesis is true (of me, at least) without any detailed empirical investigation. In a similar way we can provide a solution to the problem of armchair knowledge that is posed by my proposal about colour concepts.

Conclusion

I began by reviewing Aunty's own argument for the language of thought and have then responded to the 'no need to descend' objection, addressed the worry about eliminativism, and proposed a solution to the problem of armchair knowledge.

It is plausible that the worry about eliminativism provides some of the motivation for the 'no need to descend' objection. The downward inference from the personal to the subpersonal level seems particularly problematic to the extent that it poses a threat of eliminativism. I have argued that we do not have to choose between, on the one hand, acknowledging that in a disobliging world we should be rationally bound to abandon our folk psychological practice and, on the other hand, rejecting all such downward inferences. There is a third option. We should be open to the possibility of conceptual revision in the light of empirical discoveries. If I am right about this then the motivation for the 'no need to descend' objection is reduced. But we should retain the idea that, for conceptualised thought, there is a personal-level requirement of awareness of logical form in addition to the subpersonal-level requirements of the presence of tacit knowledge of rules of inference and the truth of the LOT hypothesis.

The worry about eliminativism can be addressed. But it points towards a genuine and serious problem, namely, that Aunty's argument seems to offer a route to armchair knowledge about the workings of the information-processing machinery inside our heads. Furthermore, our solution to the problem of armchair knowledge depends crucially on a concession that was made to the worry about eliminativism. We respect the intuition of non-negotiability concerning our engagement in folk psychological practice by acknowledging a pair of constraints on conceptual negotiation. A revised conception of a thinking being should be a conception under which we fall and it should rationally sustain as much as possible of our folk psychological practice. This pair of constraints reflects the idea that, along with a theoretical component, the concept of a thinking being has an exemplar component: we, at least, are thinking beings. Without the proposal that the concept has relatively independent exemplar and theoretical components, the solution to the problem of armchair knowledge would not work. The solution depends on the fact that acceptance of a background assumption about conceptual coherence cannot be rationally combined with a particular empirical doubt. And it is only because of the independence of its two components that the concept of a thinking being lives at empirical risk of being torn apart.

References

- Block, N. 1981: Psychologism and behaviorism. *Philosophical Review*, 90, 5–43.
- Block, N. 1995: The mind as the software of the brain. In D.N. Osherson and E.E. Smith (eds), *An Invitation to Cognitive Science, Volume 3: Thinking*. Cambridge, MA: MIT Press, 377–425.
- Brewer, B. 1995: Compulsion by reason. *Proceedings of the Aristotelian Society, Supplementary Volume* 64, 237–53.
- Brewer, B. 1996: Internalism and perceptual knowledge. *European Journal of Philosophy*, 4, 259–75.
- Brewer, B. 1999: *Perception and Reason*. Oxford: Oxford University Press.
- Chomsky, N. 1965: *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Davies, M. 1987: Tacit knowledge and semantic theory. Can a five per cent difference matter? *Mind*, 96, 441–62.
- Davies, M. 1992: 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, 235–71.
- Davies, M. 1995: Two notions of implicit rules. In J.E. Tomberlin (ed.), *Philosophical Perspectives 9: AI, Connectionism and Philosophical Psychology*. Atascadero, CA: Ridgeview Publishing Company, 153–83.
- 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, 321–61.
- Davies, M. 2000a: Persons and their underpinnings. *Philosophical Explorations*, 3, 43–62.
- Davies, M. 2000b: Externalism and armchair knowledge. In P. Boghossian and C. Peacocke (eds), *New Essays on the A Priori*. Oxford: Oxford University Press, 384–414.
- Evans, G. 1981: Semantic theory and tacit knowledge. In S. Holtzman and C. Leich (eds), *Wittgenstein: To Follow a Rule*. London: Routledge and Kegan Paul, 118–37. Reprinted in *Collected Papers*. Oxford: Oxford University Press, 1985.
- Evans, G. 1982: *The Varieties of Reference*. Oxford: Oxford University Press.
- Fodor, J. 1985: Fodor's guide to mental representation. *Mind*, 94, 55–97. Reprinted in *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press, 1990.
- Fodor, J. 1987: *Psychosemantics*. Cambridge, MA: MIT Press.
- Jackson, F.C. 1987: *Conditionals*. Oxford: Blackwell Publishers.
- Lewis, D. 1997: Naming the colours. *Australasian Journal of Philosophy*, 75, 323–42.
- Peacocke, C. 1983: *Sense and Content*. Oxford: Oxford University Press.
- Peacocke, C. 1990: Content and norms in a natural world. In E. Villanueva (ed.), *Information, Semantics and Epistemology*. Oxford: Basil Blackwell, 57–76.
- Peacocke, C. 1992: *A Study of Concepts*. Cambridge, MA: MIT Press.
- Polanyi, M. 1967: *The Tacit Dimension*. London: Routledge and Kegan Paul.
- Ramsey, W., Stich, S. and Garon, J. 1990: Connectionism, eliminativism and the future of folk psychology. In J.E. Tomberlin (ed.), *Philosophical Perspectives 4: Action Theory and Philosophy of Mind*. Atascadero, CA: Ridgeview Publishing Company, 499–533.

- Rey, G. 1995: A not 'merely empirical' argument for a language of thought. In J.E. Tomberlin (ed.), *Philosophical Perspectives 9: AI, Connectionism and Philosophical Psychology*. Atascadero, CA: Ridgeview Publishing Company, 201–22.
- Wright, C. 1985: Facts and certainty. *Proceedings of the British Academy*, 71, 429–72.