

KNOWING WHAT THOUGHT REQUIRES

Part 1: Architecturalism and the Language of Thought

MARTIN DAVIES

A Priori

www.apriori.canterbury.ac.nz

Volume 1, No. 1, pp. 111-120

January 2001

©*Martin Davies*

Acknowledgements

The material that follows is the slightly expanded text of three lectures given while I was an Erskine Fellow in the Department of Philosophy and Religious Studies at the University of Canterbury, Christchurch, New Zealand. I am very grateful to the members of the Department for inviting me to spend a month at Canterbury and for providing the opportunity to develop and present my work in a systematic way and at some length. I learned much from their questions, comments and conversations.

Lectures 1 and 2 draw on two forthcoming papers: ‘Aunty’s argument and armchair knowledge’, to appear J.M. Larrazabal and L.A. Perez-Miranda (eds), *Language, Knowledge and Representation* (Dordrecht: Kluwer Academic Publishers), and ‘Externalism, self-knowledge and transmission of warrant’, to appear in M.J. Frápolli and E. Romero (eds), *Meaning, Basic Self-Knowledge, and Mind: Essays on Tyler Burge* (Stanford: CSLI Publications). Lecture 3 develops some material first presented in a talk at the Australian National University in September 1998.

I have been thinking about these issues concerning architecturalism, externalism, and epistemic warrant for about ten years. During that time, I have been helped by many friends and colleagues including Antonia Barke, Helen Beebee, Paul Boghossian, Bill Brewer, Jessica Brown, Graeme Forbes, Jay Garfield, Mark Greenberg, Richard Holton, Frank Jackson, Wolfgang Künne, Rae Langton, David Lewis, Brian Loar, Kirk Ludwig, Michael McKinsey, Brian McLaughlin, Michael Martin, Joseph Owens, Christopher Peacocke, Paul Pietroski, Diana Raffman, Sarah Sawyer, Stephen Schiffer, Nigel Shardlow, Ernest Sosa, Daniel Stoljar, Tom Stoneham, Mark Textor and Crispin Wright. I am grateful to them all.

Lecture 1: Architecturalism and the Language of Thought

Introduction: In the armchair, down and out

In these three lectures, I shall be developing and defending the view that being a thinking and experiencing person depends on being embodied and embedded in the right way. The claim about embodiment, that thought requires a particular kind of internal cognitive machinery, is an *architecturalist* claim. The claim about being embedded, that there are requirements that our environment must meet if we are to have experiences and thoughts with certain contents, is an *externalist* claim. I defend these claims with philosophical arguments of a relatively a priori kind – arguments advanced from the armchair. These architecturalist and externalist arguments are of some independent interest. I shall consider them in the first and second lectures. But I am mainly concerned with a serious epistemological problem that they pose.

The problem becomes visible when the arguments are combined with an independently plausible claim about self-knowledge: we can know about our own thoughts and their contents from the armchair. My first-personal knowledge that I am thinking and what I am thinking does not depend for its status as knowledge on my conducting any detailed empirical investigation either of the information processing going on inside my head or of the physical and social environment in which I am situated. I am able to know from the armchair that I am a thinking being and that I think many particular things. But if the architecturalist and externalist arguments that I advance are correct and if, as I suppose, philosophical arguments yield knowledge, then there is more that I can know from the armchair. I can know, not only that I am a thinking being, but also that thought requires a particular kind of internal cognitive machinery. I can know, not only that I have thoughts with certain particular contents, but also that having those thoughts imposes requirements on my environment. But then, it seems, I must be able to put these pieces of

knowledge together. If I can know from the armchair that I am a thinking being and also that thought requires a particular kind of internal cognitive machinery then I seem to have a route to armchair knowledge about my cognitive architecture. If I can know from the armchair that I have thoughts with certain particular contents and also that having those thoughts imposes requirements on my environment then I seem to have a route to armchair knowledge about the external world. These consequences are seriously counterintuitive. How can I know about my cognitive architecture without this knowledge depending for its status as knowledge on my conducting some detailed empirical investigation of the information processing going on inside my head? How can I know about the external world without this knowledge depending for its status as knowledge on my conducting some detailed empirical investigation of the physical and social environment in which I am situated?

Architecturalist and externalist arguments generate two instances of *the problem of armchair knowledge*. When these arguments are combined with a claim about self-knowledge they seem to yield deeply implausible consequences about what it is possible to know from the armchair. Given the plausibility of the claim about knowledge of our own thoughts, the problem of armchair knowledge is naturally regarded as casting doubt on the arguments that generate it. The moral that many will draw is that armchair philosophical theorising cannot 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. But I shall be taking a different approach.

In my view, philosophical theorising, conducted in the armchair, can indeed support both conditional claims that link the personal level of folk psychology with the subpersonal level of information-processing mechanisms and conditional claims that link mind and world. In the armchair, we can proceed both down and out, to know what thought requires. But I also want to

maintain the plausible claim about first-personal knowledge of our thoughts and their contents. Deeply implausible consequences seem to follow. So I must either deny that they 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). But as a general strategy it does not appeal. It surely is implausible that we should be able to know substantive facts about our cognitive architecture without empirically investigating the information-processing machinery in our heads. So I shall be taking the second option.

In the armchair, I can know what thought requires. In the armchair, I can know about my thoughts and their contents. But I cannot know, purely by armchair reflection, that the conditions that thought requires are conditions that actually obtain. In general, from the facts that I can have armchair knowledge of a conditional (if A then B), and that I can have armchair knowledge of the antecedent of the conditional (A), it does not follow that I can have armchair knowledge of the consequent of the conditional (B). In my view, then, the solution to the problem of armchair knowledge lies in limitations on our ability to gain knowledge by inference from things that we already know. Sometimes, the epistemic warrant or justification that we have for believing the premises of an argument is not transmitted to the conclusion of the argument, even though the argument is palpably valid.

Placing limitations on the transmission of epistemic warrant from premises to conclusion in palpably valid arguments may strike you as an extreme measure. Knowledge by inference is surely a vital component in our epistemic practices. So it may seem to you much more promising to reject the arguments that generate instances of the problem of armchair knowledge or, if any of the arguments withstand critical assessment, simply to bite the bullet: more can be known from the armchair than might initially have been supposed. In my view, you would be right to insist that more has to be done to show that my approach is well motivated. In the third lecture, I shall address this issue in two ways. First, I shall show that instances of the problem of

armchair knowledge, or closely related problems about transmission of epistemic warrant, are relatively widespread. It would not be right to suppose that they are only generated by a couple of idiosyncratic philosophical arguments. Second, I shall show that the limitations on transmission of warrant are far from being ad hoc. Failure of transmission of epistemic warrant is the analogue, within the thought of a single subject, of the dialectical phenomenon of begging the question.

1. Aunty's own argument for the language of thought

About twelve years ago, 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 lecture, I briefly review Aunty's argument and then respond to an objection, address a worry, and draw attention to a serious problem.⁵

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 will invariably 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.

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.

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 taken it as a background assumption that thoughts are underpinned by physical configurations; but we have *not* taken it as a background assumption that those physical configurations 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, because of the quite general connection between tacit knowledge of rules and syntactically structured representations, the second step of Aunty’s argument

tells us that any physical configuration that figures as an input to component processors embodying tacit knowledge of rules of inference must be syntactically structured.

Consider now the 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 By the first step of Aunty’s argument, the thinker meets the conditions for tacit knowledge of the corresponding rule of inference. Suppose, too, that the thought, as it occurs on this occasion, is apt to figure as a premise in an inference of this form. Then, by the second step of Aunty’s argument, 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.

I said that I would review Aunty’s argument and then respond to an objection, address a worry, and draw attention to a serious problem. In this section I have reviewed the argument. I now turn to the objection.

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. Some-

one 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 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.

I said that, following a brief review of Aunty's argument (Section 1), I would respond to an objection, address a worry, and draw attention to a serious problem. In this section I have responded to the 'no need to descend' objection. In the next section I shall address the worry.

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

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 pros-

pect 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 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 conditions and in a disobliging world the two components lead to contradictory verdicts on cases. I claim that the option of being open to the possibility of conceptual revision in the light of empirical discoveries is preferable to the more extreme options of standing ready to perform an eliminativist *modus tollens* or else 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 (Section 2) and addressed the worry about eliminativism (Section 3). But there is still a serious problem for Aunty's argument.

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. The problem of armchair knowledge is certainly serious. It may well 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.

Solving the problem of armchair knowledge is my project for the third lecture. 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.¹⁹ But before turning to the solution to the problem, I want to present another case in which the premises of a simple *modus ponens* inference can be known from the armchair, yet knowledge of the conclusion requires an investigative methodology.