In P. Carruthers and J. Boucher (eds), *Language and Thought*
(Cambridge: Cambridge University Press, 1998), 226–47.

Language, Thought, and the Language of Thought (Aunty's Own Argument Revisited)*

MARTIN DAVIES

1. Introduction

In this chapter, I shall be examining an argument for the language of thought hypothesis – an argument which, in earlier work (Davies, 1992; see also 1991), I have called 'Aunty's own argument for the language of thought'. That will be the business of Sections 2-5. In the final section, I shall briefly mention some points of contact between this argument for the language of thought (LOT) hypothesis and the hypothesis that is the topic of Peter Carruthers's book, *Language, Thought and Consciousness*, which I shall call the thinking in natural language (TNL) hypothesis. Before beginning on Aunty's own argument, however, I shall briefly present a framework for organising questions about the relative priority of thought and language.

1.1 Orders of priority

Should questions in the theory of thought – questions about intentionality, beliefs and concept possession, for example – be approached directly or, instead, indirectly via questions about language? Suppose that Kylie believes that kangaroos seldom kick, and expresses this thought in the English sentence: 'Kangaroos seldom kick'. Which takes priority, the meaning of the English sentence or the content of Kylie's thought?

A claim of priority is the converse of a claim of one-way dependence: X enjoys priority over Y if Y depends on X but X does not depend on Y. So, any question of the relative priority of X and Y has four possible answers: (i) X has priority; (ii) Y has priority; (iii) X and Y are mutually dependent (interdependent); (iv) X and Y are independent. But the question of the relative priority of thought and language is unclear until the relevant kind of priority has been specified. I suggest that it is useful to distinguish three kinds of priority question: *ontological*, *epistemological*, and *analytical* (see Avramides, 1989, for a similar distinction).

To say that thought enjoys *ontological priority* over language is to say that language is ontologically dependent on thought, while thought is not so dependent on language. That is, there cannot be language without thought, but there can be thought without language. To say that thought enjoys *epistemological priority* over language is to say that the route to knowledge about language (specifically, about linguistic meaning) goes via knowledge about thought (specifically, about the contents of thought), while knowledge about thought can be had without going via knowledge about language.

Donald Davidson, for example, is a philosopher who would deny both these priority claims. As for ontological priority, he argues (Davidson, 1975) that there cannot be thought without language: in order to have thoughts (specifically, beliefs), a creature must be a member of a language community, and an interpreter of the speech of others. As for epistemological priority, he argues (Davidson, 1974) that it is not possible to find out in detail what a person believes without interpreting that person's speech.

* Early versions of some sections of this chapter have been included in talks given at the University of Michigan and the University of Tasmania, at a workshop on the personal/subpersonal distinction held at the University of Stirling, and at a conference of the European Society for Philosophy and Psychology held in Barcelona. I am grateful to the audiences in all these places, and to Peter Carruthers for several very helpful comments on later versions.

Our third kind of priority, *analytical priority*, is priority in the order of philosophical analysis or elucidation. To say that X is analytically prior to Y is to say that key notions in the study of Y can be analysed or elucidated in terms of key notions in the study of X, while the analysis or elucidation of the X notions does not have to advert to the Y notions. If we fix on the notion of thought content, or intentionality, as a key notion in the study of thought, and the notion of linguistic meaning as a key notion in the study of language, then the four possible positions on the relative analytical priority of thought and language can be sketched as follows.

(i) Priority for thought: This is the view that a philosophical account of the content of thoughts can be given without essential appeal to language, and that the notion of linguistic meaning can then be analysed or elucidated in terms of the thoughts that language is used to express. Paul Grice's programme in the philosophy of language (Grice, 1989; Schiffer, 1972) was aimed, not merely at elucidation, but, more boldly, at an analysis of public language meaning in terms of the beliefs and intentions of language users. Grice did not, himself, offer any elucidatory account of the intentionality of mental states. But recent work in the philosophy of mind has brought forward several proposals for explaining the intentionality of mental states without appeal to linguistic meaning, including accounts in terms of causal covariation, of teleology, and of functional role. So, we could imagine an elucidatory programme coupling one of these accounts of thought content with a Gricean analysis of linguistic meaning in terms of mental notions. In fact, it is now widely agreed that the Gricean analytical programme cannot be carried through (Schiffer, 1987). But, even if this is right, it need not rule out the possibility that thought enjoys analytical priority over language, provided that there is some other way of elucidating (if not analysing) the notion of linguistic meaning in terms of thought content. Such an elucidation might follow the Gricean model by adverting very directly to the communicative use of language. But, in principle, it might equally proceed in two stages, first introducing a notion of idiolect meaning, and then explaining the idea of a public communicative practice in terms of shared, or overlapping, idiolects.

(ii) Priority for language: On this, opposite, view, an account of linguistic meaning can be given without bringing in the intentionality of thoughts, and what a person's thoughts are about can then be analysed in terms of the use of language. This view can be found in Michael Dummett's work (Dummett, 1973, 1991, 1993). If a theorist attempts to give a substantive account of linguistic meaning in accordance with this view then the resources that can be invoked are seriously limited, since the account cannot presume upon everyday psychological notions such as belief and intention. Because of this, it would not be surprising to find hints of behaviourism in work that is influenced by this view.

(iii) No priority – Interdependence: This – the first of two possible 'no priority' positions – is the view that there is no way of giving an account of either intentionality or linguistic meaning without bringing in the other member of the pair. The two notions have to be explained together. This is the view of Davidson (1984), who thus maintains a combined ontological, epistemological and analytical no-priority position. These three no-priority claims go together quite naturally, but it is important to note that they are separable, and presumably logically independent, claims. The analytical no-priority claim is not entailed by the ontological no-priority claim, nor by the epistemological no-priority claim, nor by the two together.

(iv) No priority – Independence: This is the view that the notions of thought content and of linguistic meaning are unrelated. This position might be defended if a language is considered as an abstract entity, composed of a set of expressions together with a function

that assigns a value to each expression (a proposition to each sentence, for example). On such a conception, meaning is a purely formal notion. But for the notion of linguistic meaning as it applies to a natural language in use, this fourth view is implausible.

The point (mentioned in (iii)) that ontological, epistemological, and analytic priority claims are independent of each other is a quite general one. It would be consistent to maintain, for example, that thought enjoys ontological priority over language (that there can be thought without language, but not language without thought), while denying that thought comes before language in the order of philosophical elucidation. Equally, it would be consistent to deny that thought enjoys ontological priority over language – insisting, instead, that there can be no thought without language – while yet maintaining that thought comes first in the order of philosophical elucidation.

Indeed, it seems that these combinations remain consistent even if we consider ontological priority, or no-priority, claims that are supposed to be established by more or less purely philosophical arguments. Thus, for example, it would seem to be consistent to combine the claim that it is a conceptual truth that there can be thought without language (conceptually based ontological priority claim) with the claim that thought does not come first in the order of philosophical elucidation (analytical no-priority claim). Similarly, it would seem to be consistent to say both that it is a conceptual truth that there cannot be thought without language (conceptually based ontological no-priority claim) and that thought comes first in the order of philosophical elucidation (analytical priority claim). Such combinations of views may, though, be unattractive and difficult to motivate. Suppose, for example, that someone proposes a specifically Gricean version of the analytical priority of thought over language, according to which linguistic meaning involves a complex structure of beliefs and intentions in a population of language users. Then, an argument for the ontological no-priority claim (in particular, for the claim that there could not be thought without language) would have to show that there could not be any beliefs or intentions at all unless there was this complex structure of beliefs about beliefs about intentions; and it is far from easy to see how that could be shown.

The argument for the language of thought, to which we now turn, makes use of ideas that emerge within a framework that accords analytical priority to thought over language (Evans, 1982; Peacocke, 1992). While those ideas do not involve any specifically Gricean commitments, it would be fair to say that an assumption of ontological priority of thought over language is also in the background.

2. Aunty's own argument revisited

Jerry Fodor's Aunty 'speaks with the voice of the Establishment' (Fodor, 1987, p. 135) and is represented by Fodor as someone who resists the LOT hypothesis, preferring, perhaps, a connectionist picture of the mind's operations (*ibid*. p. 139). Aunty, as I imagine her to be, is a neo-Fregean who also maintains a proper regard for the work of Wittgenstein. The neo-Fregean framework provides – ironically enough, given Fodor's own presentation of Aunty – the resources for a relatively non-empirical argument for the LOT hypothesis. Thus, Aunty's own argument shows that the LOT hypothesis is derivable from what might be the best available philosophical account of what it is to be a thinking being. On the other hand, Aunty's residual Wittgensteinian tendencies oblige her to check possible reasons for being sceptical about the very idea of a language of thought.

2.1 Two possible reasons for scepticism

One idea that would lead to scepticism is that the LOT hypothesis is bound to be regressive, involving either a regress of languages or a regress of interpreters (cf.

Wittgenstein, *Blue Book* (1969), p. 3). It is, by now, a familiar point that the LOT hypothesis does not involve any such regress because sentences of the LOT are not presented to the thinking subject, nor to an inner homunculus, as syntactic objects standing in need of interpretation. But perhaps we can add one remark.

When I hear a sentence in a language that I understand, I do not hear the sentence as a phonological object standing in need of interpretation; rather, I hear the sentence as having a meaning. We might say that I hear the meaning clothed in phonology. So, now someone might suggest something similar, in the case of the LOT. The suggestion would be that, while LOT sentences are not presented to the thinker as syntactic vehicles to which a meaning has to be assigned, still, in conscious thought at least, LOT sentences are presented to the thinking subject, but presented as interpreted. Thus, the non-semantic properties of LOT sentences can provide a phenomenal clothing for the contents of conscious thoughts. While it may be possible to make something of this suggestion, it is no part of the LOT hypothesis as I am conceiving it here. The LOT hypothesis as it figures in Aunty's own argument is a hypothesis about cognitive processing machinery; it is pitched at the subpersonal, rather than the personal, level. For that reason, it makes no contribution to the vitally important topic of conscious thought.

A second idea that might lead to scepticism about the very idea of a language of thought would be inspired by a familiar passage from *Zettel* (1981, p. 106): 'But why should the *system* continue further in the direction of the centre? Why should this order not proceed, so to speak, out of chaos?' We should not simply slide from the fact that psychological descriptions have a certain structure – an articulation into specific belief attributions, intention attributions, as so on – to the assumption that there is a matching articulation in the physiological structure of the brain. But there are two things to be said here. One is that the LOT hypothesis, while it is not pitched at the personal level of description, is not pitched at the physiological or neuro-anatomical level either. It is a hypothesis within information-processing psychology, constrained by the physiological facts, but nevertheless at some degree of abstraction from them. The second thing to be said is that Aunty's own argument is an argument. There may, of course, be something wrong with the argument. But what Aunty proposes is not simply to project the structure of commonsense (folk) psychological descriptions onto the subpersonal level information-processing substrate.

2.2 The first step in Aunty's own argument: Systematicity and syntax

Aunty's own argument makes use of the notion of a tacitly known (or implicit) rule, where that notion in cashed out in terms of a certain kind of systematicity of causal processes. (See Davies, 1987, for some details omitted here; and Davies, 1995, for the notion of implicit rule applied to connectionist networks.) This notion of systematicity of process can be shown to require a certain structure in the inputs to the process, a structure that turns out to meet the minimal conditions for being syntactic structure. The first step in the argument is thus to establish a connection between implicit rules and syntactically structured representations. The notions that are involved in this first step – notions of causally systematic process, tacitly known rule, and syntactically structured input state – can be explained quite independently of any consideration of the LOT hypothesis.

The causal processes that are 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 described

semantically. Thus, for example, it might be that whenever the input configuration represents a letter string beginning with 'b' the output configuration represents a pronunciation beginning with /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, for the sense of tacitly known, or implicit, rule that is in play here, 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. Nor is it sufficient that this conformity to the rule should be non-accidental, holding good in nearby counterfactual situations as well as in the actual situation.

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.

Suppose that our transition mediating mechanism meets this condition, and so embodies tacit knowledge of this spelling-sound rule. Then the various input configurations that represent letter strings beginning with 'b' need to share some physical 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 letter strings beginning with '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 (Fodor, 1987, pp. 16-21). Thus, quite independently of any consideration of the LOT hypothesis, we have the result that where transition mediating mechanisms embody tacit knowledge of rules there we find syntactically structured input representations.

2.3 The second step in Aunty's own argument: Inferences and their forms

Aunty's own argument also makes use of a certain notion of inferential transitions between thoughts. (In earlier work, Davies, 1991, 1992, I developed this step of the argument in two slightly different ways, drawing in turn on Evans, 1982, and on Peacocke, 1992. Here I follow the version that uses the notion of possession conditions in Peacocke, 1992, chapter 1. One of Peacocke's proposals is that concepts can actually be individuated by their possession conditions, where those conditions are specified in terms of something like a functional (inferential) role. We should note, however, that Peacocke has subsequently changed his views somewhat; see his 1998a.) The key idea is that being able to think particular types of thought (that is, possessing particular concepts) involves a thinker in commitments to particular forms of inference. According to this idea, what is required of a subject is not just commitment to each of a number of inferences that happen to instantiate a particular form. Rather, the commitment is to accept (indeed, to perform) these inferences *in virtue of their form* (Peacocke, 1992, p. 6). The form of the inferences should figure in the causal explanation of the thinker's performing those inferences. One way to cash this out, without requiring the thinking subject to be able to specify the form of the inferences, nor to be able to offer an explicit account of the form as part of his or her reason for making the inferential transitions (*ibid*. p. 135), would be to require that the thinker meet the conditions for tacit knowledge of the inferential rule.

In order to see the consequences of this proposed way of cashing out the requirement, suppose that an occurrent thought involves the tokening of a specific physical configuration. (We assume, that is, intentional realism; see Fodor, 1987, p. 135. See also

the discussion of propositional modularity in Ramsey, Stich and Garon, 1990.) If we apply the idea of a tacitly known rule to the case of causal transitions between these physical configurations, then we shall arrive at the conclusion that the physical configuration whose tokening is the information processing level correlate of a person's thinking a particular thought is syntactically structured. In short, we shall arrive at a version of the LOT hypothesis.

We should note two points about this second step in Aunty's own argument. One point is that it only takes us from intentional realism or propositional modularity to the LOT hypothesis. Properly speaking, Aunty needs to offer an argument to support the assumption of intentional realism. The second point is that Aunty's own argument involves a transition from the personal level to the subpersonal level of cognitive machinery – from a thinking person finding inferences compelling in virtue of their form to a requirement on causal mechanisms in the cognitive machinery, a requirement of causal systematicity of transitions that leads to a requirement of syntactic structure in representations.

In the next three sections, we address three problems for Aunty's own argument. First, the argument seems to present an invitation to eliminativism (Section 3). Second, the argument seems to offer a non-empirical route to substantive knowledge about the world inside our skulls (Section 4). Third, the argument may seem to be undermined by the very fact that it moves from the personal to the subpersonal level (Section 5).

3. Eliminativism and conceptual negotiation

Aunty's own argument uncovers a necessary condition for a physical being to be a thinking person, and the necessary condition concerns internal cognitive architecture: a thinking being must be an LOT being. This cognitive architectural condition evidently goes beyond facts about behaviour. Thus, given any physical being whose behaviour prima facie warrants the attribution to it of beliefs and other attitudes, in accordance with the intentional stance (Dennett, 1987), it is an epistemic possibility that the being does not meet the condition on internal cognitive architecture. So, Aunty's own argument appears to present an invitation to eliminativism. If we turn out not to be LOT beings, then we also turn out not to be thinking persons.

3.1 Theoretical options in a disobliging world

In order to see what is at issue here, suppose, for a moment, that developments in the scientific investigation of the mind – whether in cognitive psychology or in neuroscience – were to show that the cognitive architectural condition was not, in fact, met; in short, that we are not LOT beings. Then there would be a number of theoretical options available. One option would be to conclude that some of the pieces of philosophical theory drawn on in Aunty's own argument are wrong. Another option – the opposite extreme – would be to abandon wholesale our folk psychological practice of describing, interpreting, and explaining what people do as acting for reasons that are based on beliefs, wants, hopes, fears, and the rest.

If we consider only these two theoretical options, then the thought that the second option is not genuinely available to us – that our engagement in ordinary folk psychological practice is philosophically non-negotiable – may seem to constitute a powerful objection to Aunty's own argument. If that objection is a good one, then it does not apply to Aunty's own argument alone. Rather, it would apply equally to any philosophical argument that appears to uncover substantive cognitive architectural necessary conditions for being a thinking person. So, the question that we need to ask is

whether one can argue from the non-negotiability of our engagement in folk psychological practice to the incorrectness of all such *architecturalist* arguments.

One problem with this putative line of argument is that it depends on overlooking a third theoretical option, namely, that we might maintain our folk psychological practice even though many of the claims made in folk psychological descriptions, interpretations, and explanations were false. But even setting that problem aside, there is a worry about the idea of a blanket rejection of all architecturalist arguments. For the competing piece of conceptual analysis that would be suggested by that rejection is itself arguably out of line with our intuitive judgements about which physical beings are thinking persons.

What the blanket rejection of all arguments that uncover cognitive architectural commitments suggests is that an analysis of the concept of a thinking person should impose no necessary conditions at all on internal cognitive architecture and, indeed, no necessary conditions that go beyond behaviour. This entails that if two physical beings are behavioural (or, perhaps better: trajectorial) duplicates in actual, and nearby counterfactual, situations, then either both are thinking persons or neither is. But that doctrine is revealed as being out of line with our intuitions when we consider imaginary examples of physical beings that produce the right behaviour by way of unusual internal architectures, such as the string-searching machine of Block (1981) or the Martian marionette of Peacocke (1983). The string-searching machine, which stores a finite but massive collection of interpretable sequences of behaviour, can *ex hypothesi* meet any behavioural requirements for being a thinking person. But, as Block remarks (1990, p. 252), 'it has the intelligence of a jukebox'.

The two options of, on the one hand, rejecting all architecturalist arguments and, on the other hand, abandoning our folk psychological practice if things turn out badly, do not exhaust the options. In between, there lies the possibility of conceptual negotiation. In order to see how this possibility would work, suppose that the pieces of philosophical theory that are drawn on in Aunty's own argument do correctly elaborate and precisify our current conception of a thinking person and that the argument correctly uncovers the commitments of that conception. Then imagine that things turn out badly – that we turn out not to be LOT beings. Evidently, in those circumstances, we ourselves would not fall under (the best elaboration and precisification of) our current conception of a thinking person. But we might still be able rationally to sustain the greater part of our folk psychological practice if we could negotiate our way to a new, revised, conception of what it is to be a thinking person. The details of the negotiations would depend on the particular ways in which things turned out badly. They would also depend on the philosophical theories connecting those empirical discoveries with our current conceptions of folk psychological phenomena.

3.2 The appeal to consciousness

The upshot of our discussion so far in this section is intended to be that it is no objection to an architecturalist argument, such as Aunty's own argument, that it presents the possibility of an eliminativist modus tollens. But it might be said that there is something unsatisfactory about the way that we have dealt with the presumed non-negotiability of folk psychological practice. Our strategy (two paragraphs back) was to suggest that, if what is wanted is a guarantee that those who engage in the practice really are thinking persons, then the price to be paid is commitment to a counter-intuitive doctrine about trajectorial duplicates. In response to this strategy, it might be said that there is something that can be known about persons, at least in the case of the first person singular, which goes beyond behaviour yet has nothing to do with internal cognitive architecture. So, it

may be said, it is possible to reject all architecturalist arguments without paying the price of commitment to the counter-intuitive doctrine about trajectorial duplicates. For it is open to someone to maintain that an account of thinking persons should impose no necessary conditions that go beyond behaviour *plus consciousness*.

There is more than one way to develop this suggestion, and some of the ways do not seem satisfactory at all. Thus suppose, for example, that we focus on the notion of consciousness that applies to itches, pains and tickles. Though the issues are complex, it is difficult to be convinced that we move closer to a thinking person by adding bare sensations to a string-searching machine or a Martian marionette. But suppose that we consider, not sensations, but conscious thought. Then the suggestion may be that, by introspection, I can know that I think many things, and more generally can know that I am a thinking being. So, an account of thinking persons may go beyond behaviour by adverting to what can be known by introspection, yet without taking on any cognitive architectural commitments.

There are two points to note about this suggestion. One point is that it cannot really underwrite the non-negotiability of folk psychological practice, since it is restricted to the first person singular. The second point is that, in the face of an architecturalist argument, such as Aunty's own argument, the suggestion is apt to seem question-begging. It may be true that thinking persons can know by introspection that they think many things, and can know that they are indeed thinking beings. But, according to Aunty's own argument, if we turn out not to be LOT beings then we shall not be thinking beings. In those circumstances, we should not be able to know, whether by introspection or any other way, that we are beings who think.

The suggestion emerging from the appeal to consciousness seems to be question-begging. But, in fact, the idea of first personal introspective knowledge of our thoughts leads to a serious problem for Aunty's own argument, and the solution to the problem involves a further concession to the intuition of non-negotiability. This is the topic for the next section.

4. The problem of armchair knowledge

The problem to be considered in this section arises when we consider arguments of the following general form (MC):

(1) I have mental property M.
(2) If I have mental property M then I meet condition C.
Therefore:
(3) I meet condition C.

in cases where premise (2) is justified by an architecturalist argument, such as Aunty's own argument. Specifically, the problematic argument is LOT(MC):

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.

Let us say that a thesis of *first person authority* about a mental property, M, is a thesis to the effect that we have a distinctively first personal and specially authoritative way of knowing that we ourselves have property M, when we do have it, without needing to conduct any detailed empirical investigation of the world outside or within. If we assume a thesis of first person authority about the property of being a thinking being – in line with the suggestion at the end of Section 3 – then the problem is clear.

I do not need to engage in any detailed empirical investigation of my internal cognitive architecture to know premise LOT(1); I know with first person authority that I think many things. Nor, do I need to engage in empirical research in order to know premise LOT(2); it is underwritten by a philosophical argument. It is obvious that the conclusion LOT(3) follows from these two premises. Yet, even supposing that the conclusion is true, it is massively implausible that its truth can be known from the armchair. Questions about internal cognitive architecture – about whether we are LOT beings, for example – cannot be settled without major programmes of empirical research. This is the problem of armchair knowledge. Aunty's own argument, in combination with a thesis of first person authority, seems to offer us an unacceptably non-empirical route to knowledge of substantive empirical facts about cognitive architecture.

4.1 Limiting knowledge by inference

In recent work (Davies, 1998), I have suggested that the way to deal with this problem is to impose limitations on knowledge by inference. As a first attempt, I proposed two limitation principles:

> **First Limitation Principle:**
> Epistemic warrant cannot be transferred from A to B, even given an a priori known entailment from A to B, if the truth of B is a precondition of our warrant for A counting as a warrant.
>
> **Second Limitation Principle:**
> Epistemic warrant cannot be transferred from A to B, even given an a priori known entailment from A to B, if the truth of B is a precondition of the knower even being able to believe the proposition A.

It is the second of these that was supposed to deal with the problem of armchair knowledge posed by the LOT(MC) argument.

The principle works by blocking the transfer of epistemic warrant from premises to conclusion in that argument. Even though I can know LOT(1) and LOT(2) without rising from my armchair, I cannot, according to the Second Limitation Principle, thereby come to know LOT(3), even though it plainly follows from those premises.

According to Aunty's own argument, if I am a thinking being, then I am an LOT being. If that argument is correct then, even in order to believe that I am a thinking being, I need to be an LOT being. Aunty's own argument might be wrong, of course. But if it is right then it triggers application of the Second Limitation Principle, and epistemic warrant cannot be transferred from LOT(1) to LOT(3), even given the a priori known entailment in LOT(2). Given the Second Limitation Principle, the very argument that gives rise to the problem of armchair knowledge also provides for its solution. So, given that principle, it would be wrong to press that problem as an objection against Aunty's own argument.

The Second Limitation Principle allows us to block the unacceptably non-empirical route to knowledge of substantive empirical facts. But I provided little enough motivation for the principle, and it seems very plausible that the principle is open to counterexamples. What I propose to do here is to provide some motivation for (modified versions of) the limitation principles.

4.2 Improving the principles

The intuitive idea behind both limitation principles is something like this. In any given epistemic project, some propositions will have a presuppositional status. Suppose that the

focus of the project P is the proposition A, and that the investigation is carried out using method N. Then within project P it is presupposed, for example, that A is a hypothesis that can be coherently entertained (can be believed, doubted, confirmed, disconfirmed); and it is also presupposed that N is a method that can yield knowledge, at least with respect to A. Suppose that B is some proposition that has this presuppositional status in project P. Then P cannot itself yield knowledge that B; nor can P play an essential role in yielding knowledge that B.

The First Limitation Principle can be regarded as an attempt at codifying this intuitive idea as it relates to the presupposition about the method: as the principle is formulated, *our warrant for A counting as a warrant* stands in for the investigative method being such as to yield knowledge. The Second Limitation Principle likewise attempts to codify the idea as it relates to the presupposition about the hypothesis: as the principle is formulated, *the knower being able to believe the proposition A* stands in for the hypothesis being such as can be coherently entertained.

It is easy to construct counterexamples to the limitation principles as they were originally formulated. Consider, for example, the simple inference from the premise:

> I believe that water is wet

to the conclusion:

> Someone is able to believe something.

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

We need to tighten up the Second Limitation Principle (and the First Limitation Principle, too). To that end, let us suppose for the moment that the two basic presuppositions in an epistemic project P using method M and with target hypothesis A are:

> (i) the proposition that method M is knowledge yielding (at least with respect to A); and
>
> (ii) the proposition that there is such a proposition as the proposition A for the putative knower to entertain (that the conceptual practices on which A draws are not internally incoherent, for example).

Suppose too that other propositions become derived presuppositions in project P by being shown to follow from propositions that are already presuppositions in P by way of some circumscribed set of resources that are already in use in P. Then we might be led to an improved version of the Second Limitation Principle as follows:

> **Second Limitation Principle (revised version)**
> Epistemic warrant cannot be transferred from A to B, even given an a priori known entailment from A to B, if the truth of B can be shown *by resources that are used in the epistemic project* [e.g. the resources used to derive B from A] to be a precondition of *there being any such proposition for the knower to entertain or believe as the proposition A*.

(I shall not pause over revisions to the First Limitation Principle, since it is not directly relevant to the problem that concerns us.) But the difficulty that we now face is that the improved version of the Second Limitation Principle is no longer adequate to deal with the problem of armchair knowledge posed by Aunty's own argument.

Aunty's own argument would indeed have the consequence that if a putative knower is not an LOT being then that putative knower would not even be able to think the thought:

LOT(1) I am a thinking being.

But Aunty's own argument, which uncovers necessary conditions for being a thinking being, does not have the consequence that if the putative knower does not meet the cognitive architectural condition then there is no such proposition as the proposition LOT(1). Aunty's own argument does not reveal any internal incoherence in the very notion of a thinking being in the disobliging circumstance that the cognitive architectural necessary condition is not met.

4.3 Modifying the solution

Under what circumstances would the failure of a being X to meet a necessary condition for falling under a concept C reveal an internal incoherence within the C-conception, rather than merely revealing that X is a non-C? One circumstance would be that there are sufficient conditions for falling under the concept C, and X does meet those conditions.

More generally, a conception may involve a sufficient conditions component and a necessary conditions component. It may be – familiarly – that some objects do not meet the sufficient conditions but do meet the necessary conditions. On such objects, the conception does not pronounce one way or the other on the question, 'Is it a C?'. The possibility that I am now raising is that an object might meet the sufficient conditions but not meet the necessary conditions. On such objects, the conception produces a contradictory pronouncement; the conception is – perhaps thanks to a disobliging world – internally incoherent. Conceptual revision and negotiation are required.

One way in which this possibility could come about would involve a sufficient conditions component based on paradigm exemplars that turn out not to meet the necessary conditions component. So, suppose that the conception of a thinking being involves not only an architecturalist necessary conditions component but also an exemplar-based sufficient conditions component: you and I are to count as thinking beings. Then our meeting the architecturalist requirement for thought really would be a precondition of the internal coherence of the concept of a thinking being, of there being any such proposition as: X is a thinking being.

The upshot of this would be that, in any epistemic project in which the conception of a thinking being figures, it is a presupposition that the paradigm exemplars meet whatever necessary conditions may be built into that conception. Aunty's own argument raises the specific proposition that the paradigm exemplars are LOT beings to presuppositional status and in doing so disqualifies any proposition to the effect that some paradigm exemplar (like me) is an LOT being from being a potential recipient of epistemic warrant in a project using that argument. Thus, we may be able to solve the problem of armchair knowledge that is posed by Aunty's own argument; but the solution involves more of a concession to the non-negotiability of folk psychological practice than we envisaged in Section 3. What we are conceding here is not that those who engage in folk psychological practice are guaranteed to be thinking persons. But we are accepting that our current conception of thinking persons has no coherent use unless it applies to those – like you and me – who engage in that practice.

5. Moving from the personal to the subpersonal level

The problem of armchair knowledge, and the solution that we have proposed in Section 4, have a somewhat technical character. The worry to be addressed in this section is a

more fundamental one; namely, that Aunty's own argument is undermined by the very fact that it moves from the personal to the subpersonal level. According to this worry, a philosophical account of what it is to be a thinking being should not directly invoke any subpersonal level notions, and should not support any substantive entailments between the personal and subpersonal levels either.

My preferred conception of the inter-level relation is as *interaction without reduction*. According to that conception, what we typically find are downward entailments from the personal to the subpersonal level, but explanatory gaps when we try to reconstruct personal level notions out of subpersonal level resources. In slightly more detail, at the personal level of description we find many notions – subjective, normative – that have no place in science, and we find a distinctive kind of intelligibility. But these personal level descriptions also make use of causal notions, and the correctness of these descriptions is not indifferent to issues about subpersonal level information processing machinery (interaction – downward entailments). On the other hand, an account in information processing terms of the system that constitutes (is in the same place as) a thinking person is not adequate to supplant the original personal level description (without reduction – upward explanatory gaps).

This is not the place to give an extended defence of that conception. What I propose to do is to return to the topic of inferences that are performed in virtue of their form (Section 2.3) and elaborate the description of those inferential transitions in terms that belong very clearly to the personal level. Then I shall consider an objection to Aunty's own argument that is based on that description, and suggest that the objection depends on the idea that claims about causation at the personal level are utterly indifferent to facts about information processing machinery – a claim that we have no reason to accept.

5.1 Reasoning as a conscious, rational, knowledge-yielding activity

Suppose that Bruce believes that *A or B* and also believes that *not-A*. Then it is likely that Bruce will also believe that *B*, or will come to believe it if the question whether *B* arises. Bruce's little piece of reasoning is liable to be a conscious activity, and the transition in thought that he makes is a rational one; it is the kind of transition that could (despite the limitation principles discussed in Section 4) yield Bruce knowledge that *B* if he started out from knowledge that *A or B* and that *not-A*. What are the conditions under which we are able to regard a personal level causal transition such as this as rational and potentially knowledge-yielding?

One condition, surely, is that Bruce's first two beliefs should actually constitute a reason for believing the third thing. We would show that this condition is met by pointing out that the first two believed propositions entail the third. The argument with the first two beliefs as premises and the third belief as conclusion, instantiates a valid form. The general point here is that, in order to show how Bruce's transition at least could be a rational one, we need to conduct an investigation with an abstract subject matter: we plot the contours of the abstract space of reasons.

Our investigation of the abstract space of reasons reveals that, if Bruce believes that *A or B* and that *not-A*, then the right thing for Bruce to think in addition is that *B*. Those first two things that Bruce thinks are a reason for someone to think that *B*. But if they are to be Bruce's reason, then something more must be true: Bruce's believing, or coming to believe, the first two things must cause him to believe the third thing.

However, this is not yet sufficient for Bruce's transition to be a rational one. The problem is to connect the reason condition and the causal condition in the right way, so that it is because believing that *A or B* and that *not-A* is a reason to believe that *B* that

Bruce's believing those two things causes him to believe the third thing (see e.g. Antony, 1989, 1991; Brewer, 1995). This problem is not going to be solved here, but there are two suggestions that it would be natural to make.

The first suggestion is one that has already been mentioned (Section 2.3); namely, that Bruce should perform the transition in thought because it is of that valid form. As we have already said, this is not to require that Bruce can conceptualise or spell out what that form is; and still less is it to require that Bruce should use as an extra premise a belief that *((A or B) and not-A) entails B*. The second suggestion is that, although he need not conceptualise the form, still, in some way, Bruce should be aware of his beliefs, and the transition between them, as instantiating that form. (This second suggestion is intended to be in the spirit of Brewer, 1995.)

5.2 An objection to Aunty's own argument, and a response

Someone might now say that Aunty's own argument is undermined by our imposing these conditions at the personal level – in particular, the condition that Bruce should be aware of the form of the inference. For, if the form of the inference is already transparent to the thinker, why does it need, in addition, to be encoded in subpersonal level physical configurations?

Expressed like this, the objection misses its mark. Aunty's own argument makes use of the idea of an inferential transition being performed because it is of a certain form (where this is not to be cashed out in terms of a statement of the validity of the form being added as an extra premise to the inference). We have just suggested that, if the transition is to be rational, then the thinker should be aware of the form of the inference. But clearly, a thinker's being aware of the form of an inference is not sufficient for the form's figuring in the explanation of the thinker's performing the inference. So this recently suggested personal level condition is not one that figures directly in Aunty's own argument. Similarly, at the subpersonal level, it is not syntactic structure in physical configurations, but tacit knowledge of the rule of inference, that is supposed to cash out the requirement that the inference should be performed in virtue of its form. Syntactic structure in physical encodings is a necessary, but not a sufficient, condition for the presence of tacit knowledge.

The real question, then, is not whether Bruce's being aware of the form of his inference renders the language of thought hypothesis redundant. Rather, the question is whether the personal level requirement that Bruce should perform the transition in thought because it is of that form removes the motivation for the requirement that the cognitive machinery underpinning Bruce's inferential transitions should embody tacit knowledge of the rule of inference. It is difficult to see how the motivation for the condition on cognitive machinery would be removed by our imposing the personal level requirement, unless we could rely on a general principle to the effect that the truth of personal level claims about causal relations and the structure of causal explanations is indifferent to facts about cognitive machinery. But we have no reason to accept that general principle. (See Stone and Davies, 1993, for an example of the way in which a personal level claim about the causal order could be threatened by an empirical claim about the structure of an information processing system.)

We have been considering three problems for Aunty's own argument. The argument seems to offer an invitation to eliminativism; but in a disobliging world one option is conceptual negotiation (Section 3). The argument seems to offer a non-empirical route to substantive empirical knowledge; but that route can be blocked by limitation principles on knowledge by inference (Section 4). The argument may seem to be undermined by the

very fact that it moves from the personal to the subpersonal level; but this would be so only if causal claims at the personal level were indifferent to the causal order at the level of information processing machinery (Section 5). We turn now – finally and briefly – to the idea of thinking in natural language.

6. Thinking in natural language

There are several points of contact between Aunty's own argument for the LOT hypothesis and the thinking in natural language (TNL) hypothesis that is developed and defended by Peter Carruthers (1996, ch. 5 this volume). Carruthers sets up an opposition between the *communicative conception* and the *cognitive conception* of language, and then argues for the cognitive conception. According to the communicative conception, 'the function and purpose of natural language is to facilitate communication and *not* . . . to facilitate thinking' (1996, p. 1). According to the cognitive conception, in contrast, 'we often think *in* language, and the trains of reasoning which lead up to many of our decisions and actions will consist in sequences of natural-language sentences' (1996, p. 2).

Carruthers starts out by accepting broadly Fodorian considerations in favour of a language of thought. But Fodor's language of thought is an innate and universal language – Mentalese – whereas Carruthers defends the thesis that at least some thinking involves the thinker's natural language; in particular, that some conscious thoughts (namely, conscious propositional thoughts) are constituted by tokenings of natural language sentences. Thus, as between Fodor's LOT hypothesis and Carruthers's TNL hypothesis, 'The main focus for debate will concern *which* sentences are constitutive of our (conscious propositional) thoughts – those of Mentalese, or those of natural language' (1996, p. 39).

6.1 Imagined speech and LF representations

It is not obvious whether it is right to equate conscious propositional thoughts with thoughts that are clothed in natural language sentences. Ray Jackendoff, for example, argues that, although we 'very often experience our thought as "talking to ourselves"', still thought and language are very different phenomena (1997, p. 183). Indeed, he goes so far as to suggest that 'thought per se is *never* conscious' (*ibid*. p. 187). Many other people report that the sentences that go through their minds when they are thinking carefully and attentively seem to constitute a kind of commentary on their thinking, rather than expressing the contents of their first-order thoughts. (See also Peacocke, 1998b, for an account of conscious thinking that draws on the notion of attention, but not verbalisation.) But, whether or not it is right to make this equation, there surely is such a thing as thinking by imagining speaking. According to the TNL hypothesis, such thinking involves the activation of a PF (Phonetic Form) representation. But a PF representation is not enough to account for the fact that the imagined sentence is understood; for that, we need an LF (Logical Form) representation as well. For, while a PF representation is 'interpreted' at the 'articulatory-perceptual interface', an LF representation is 'interpreted' at the 'conceptual-intentional interface' (Chomsky, 1995, p. 219). This is to say that it is only in virtue of its relation to an LF representation that a PF representation comes to have a meaning.

The LOT hypothesis is intended to be compatible with the evident fact that people speak as well as think, that people hear meaning in natural language sentences (see Section 2.1 above), and that people can engage in imagined speech. Presumably, the LOT story about imagined speech will be somewhat similar to the TNL story, though it must

be part of the LOT story that, in addition to the activation of an LF representation, there will also be activation of an LOT sentence lying, as it were, just on the other side of the conceptual-intentional interface.

It seems plausible that a decision between these competing accounts of imagined speech will rest on detailed matters, many of which will be highly theory internal. It would be relevant to consider, for example, whether the language module was to be regarded in Fodor's way (1983) – where modules are contrasted with the central system – or in Chomsky's way (1986, p. 14) – where linguistic knowledge is embodied in a central system which nevertheless has some of the properties of Fodorian modules, 'a fact that brings the entire picture into question'. It would also be relevant to consider, in detail, the nature of the supposed conceptual-articulatory interface as that figures in Chomsky's account and, more generally, the nature of the relationship between LF representations and LOT sentences. In respect of this latter, it would be important to have a detailed account of the way that semantically non-specific LF representations map to semantically specific LOT sentences (Sperber and Wilson, 1986, ch. 9 this volume) – an account that could then be compared, in explanatory potential, with the TNL account of the semantic non-specificity of natural language sentences.

What Aunty's own argument requires is that the physical configurations whose tokenings are the information processing level correlates of a person's acts of (propositional) thinking should be syntactically structured. There is nothing in that argument to rule out the possibility that those physical configurations should turn out to be tokenings of LF representations. To that extent, Aunty's own argument, although it is presented as an argument for the LOT hypothesis, does not really differentiate between the LOT hypothesis and a version of the TNL hypothesis; namely, the hypothesis that LF representations are the subpersonal level correlates of thoughts.

On the other hand, there is a way of understanding the idea of thinking in natural language – focusing on the personal, rather than the subpersonal, level – according to which the TNL hypothesis might not measure up to the requirements imposed by Aunty's own argument.

6.2 Imagined speech and awareness of structure

Carruthers's argument for the TNL hypothesis is complex, drawing on nothing less than a theory of consciousness. But his starting point – evidence that is supposed, at least, to count as a plausibility consideration in favour of the TNL hypothesis – is furnished by introspection. Often, when we think, we find ourselves imagining, or imaging, sentences of our natural language. What this introspected evidence suggests is that, 'It is images of natural language sentences which are the primary vehicles of conscious thoughts' (1996, p. 51).

The TNL hypothesis says that conscious propositional thinking is constituted by imagings of natural language sentences rather than by tokenings of sentences in Fodor's Mentalese. As we have been understanding it (Section 2.1), Fodor's LOT hypothesis concerns the subpersonal level of cognitive machinery. In contrast, the TNL hypothesis – at least when it is conceived as a claim that can be supported by introspection – is pitched at the personal level. So, it may seem that the TNL hypothesis can offer a way of meeting the requirements of Aunty's own argument without descending to the subpersonal level of cognitive machinery. What Section 5.2 should already suggest, however, is that this idea is not right.

What Aunty's own argument requires is that the subpersonal vehicles of thought – and so, inter alia, of conscious thought – should be syntactically structured. It is not

enough that a thinker should hear or imagine *as structured* natural language sentences. To say this is not to deny the importance of the fact that someone who understands a sentence hears, or imagines, it in such a way as to be aware of the way that it is built up from familiar words and constructions. Arguably, this is vital to an epistemology of understanding. (Again, this is intended to be in the spirit of Brewer, 1995.) But from the fact that Bruce imagines the sentences '*A or B*' and '*not-A*' as structured and moves in imagination to the sentence '*B*', it does not follow that he performs that transition in virtue of its form.

To make this point clearer, let us set aside, for a moment, the idea of conscious awareness of structure, and simply consider the transition from '*A or B*' and '*not-A*' to '*B*'. Then it is clear (from Section 2.2) that the fact that the transition instantiates a form is not sufficient for the form to figure in the causal explanation of the transition. Nor is it enough that Bruce does, in the actual situation, and would, in nearby counterfactual situations, perform other transitions that instantiate the same form. Conformity to a rule – even non-accidental conformity – is not sufficient for tacit knowledge of the rule. The point in the previous paragraph is that adding in the further condition that Bruce is aware of the structure of these sentences, and of the transition between them, still does not guarantee that the form figures in the causal explanation of the transition that Bruce performs. Even if thoughts are clothed in imagined sentences, still the requirements of Aunty's own argument have to be met at the subpersonal level.

There are thus two aspects of the TNL hypothesis to be considered. One is that LF representations might serve as the subpersonal level vehicles of thought instead of sentences of Mentalese. The other is that personal level imaginings of natural language sentences might constitute conscious thoughts. From the point of view of the friend of Aunty's own argument, the second of these aspects is more problematic than the first.

References

Antony, L. (1989) Anomalous monism and the problem of explanatory force. *Philosophical Review*, 98, 153–87.

Antony, L. (1991) The causal relevance of the mental. *Mind and Language*, 6, 295–327.

Avramides, A. (1989) *Meaning and Mind: An Examination of a Gricean Account of Language*. Cambridge MA.: MIT Press.

Block, N. (1981) Psychologism and behaviorism. *Philosophical Review*, 90, 5–43.

Block, N. (1990) The computer model of the mind'. In D.N. Osherson and E.E. Smith (eds), *An Invitation to Cognitive Science, Volume 3: Thinking*. Cambridge, MA.: MIT Press, 247–89.

Brewer, B. (1995) Compulsion by reason. *Proceedings of the Aristotelian Society, Supplementary Volume* 64, 237–53.

Carruthers, P. (1996) *Language, Thought and Consciousness*. Cambridge: Cambridge University Press.

Chomsky, N. (1986) *Knowledge of Language: Its Nature, Origins and Use*. New York: Praeger.

Chomsky, N. (1995) *The Minimalist Program*. Cambridge, MA.: MIT Press.

Davidson, D. (1974) Belief and the basis of meaning. *Synthèse*, 27, 309–23. Reprinted in Davidson (1984).

Davidson, D. (1975) Thought and talk. In S. Guttenplan (ed.), *Mind and Language*. Oxford: Oxford University Press, 7–23. Reprinted in Davidson (1984).

Davidson, D. (1984) *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Davies, M. (1987) Tacit knowledge and semantic theory: Can a five per cent difference matter? *Mind*, 96, 441–62.

Davies, M. (1991) Concepts, connectionism, and the language of thought. In W. Ramsey, S. Stich and D. Rumelhart (eds.), *Philosophy and Connectionist Theory*. Hillsdale, NJ.: Lawrence Erlbaum Associates, 229–57.

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. Macdonald, B.C. Smith and C. Wright (eds), *Knowing Our Own Minds: Essays on Self-Knowledge*. Oxford: Oxford University Press).

Dennett, D.C. (1987) *The Intentional Stance*. Cambridge, MA.: MIT Press.

Dummett, M. (1973) *Frege: Philosophy of Language*. London: Duckworth.

Dummett, M. (1991) *The Logical Basis of Metaphysics*. Cambridge, MA.: Harvard University Press.

Dummett, M. (1993) *The Seas of Language*. Oxford: Oxford University Press.

Evans, G. (1982) *The Varieties of Reference*. Oxford: Oxford University Press.

Fodor, J. (1983) *The Modularity of Mind*. Cambridge, MA.: MIT Press.

Fodor, J. (1987) *Psychosemantics*. Cambridge, MA.: MIT Press.

Grice, H.P. (1989) *Studies in the Way of Words*. Cambridge MA.: Harvard University Press.

Jackendoff, R. (1997) *The Architecture of the Language Faculty*. Cambridge, MA.: MIT Press.
Peacocke, C. (1983) *Sense and Content*. Oxford: Oxford University Press.
Peacocke, C. (1992) *A Study of Concepts*. Cambridge, MA.: MIT Press.
Peacocke, C. (1998a) Implicit conceptions, understanding and rationality. In E. Villanueva (ed.), *Philosophical Issues 9*. Atascadero, CA.: Ridgeview Publishing Company.
Peacocke, C. (1998b) Conscious attitudes, attention and self-knowledge. In C. Macdonald, B.C. Smith and C. Wright (eds), *Knowing Our Own Minds: Essays on Self-Knowledge*. Oxford: Oxford University Press.
Ramsey, W., Stich, S. and Garon, J. (1990) Connectionism, eliminativism and the future of folk psychology. In J.E. Tomberlin (ed.), *Philosophical Perspectives 4: Action Theory and Philosophy of Mind*. Atascadero, CA.: Ridgeview Publishing Company, 499–533.
Schiffer, S. (1972) *Meaning*. Oxford: Oxford University Press (Second Edition, 1988).
Schiffer, S. (1987) *The Remnants of Meaning*. Cambridge MA.: MIT Press.
Sperber, D. and Wilson, D. (1986): *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.
Stone, T. and Davies, M. (1993) Cognitive neuropsychology and the philosophy of mind. *British Journal for the Philosophy of Science*, 44, 589–622.
Wittgenstein, L. (1969) *The Blue and Brown Books*. Oxford: Basil Blackwell.
Wittgenstein, L. (1981) *Zettel*. Oxford: Basil Blackwell.