Conscious experience versus conscious thought

Peter Carruthers

Are there different constraints on theories of conscious experience as against theories of conscious propositional thought? Is what is problematic or puzzling about each of these phenomena of the same, or of different, types? And to what extent is it plausible to think that either or both conscious experience and conscious thought involve some sort of self-reference? In pursuing these questions I shall also explore the prospects for a defensible form of eliminativism concerning conscious thinking, one that would leave the reality of conscious experience untouched. In the end, I shall argue that while there might be no such thing as conscious judging or conscious wanting, there *is* (or may well be) such a thing as conscious generic thinking.

1 The demands on theories of conscious experience

What needs to be explained about conscious experience is its *what it is likeness*, together with a number of surrounding puzzles. The primary demand on a theory of conscious experience is that it should explain how conscious experiences come to possess their distinctive subjective dimension, and hence explain why there should be something that it is *like* for subjects to undergo them. Arguably, a good theory should also explain the distinction between conscious and *un*conscious perceptual states, accounting for the fact that the latter *aren't* conscious.[1] It should explain how we can come to form purely recognitional concepts for our conscious experiences.[2] And a successful theory ought also

[1] Consider, for example, the kinds of visual percepts that one finds in blindsight (Weiskrantz, 1997), or in the online guidance of movement, if a 'two systems' account of vision is correct (Milner and Goodale, 1995; Jacob and Jeannerod, 2003).

[2] Such concepts are arguably at the bottom of inverted-qualia and zombie-style thought experiments. Possessing such concepts, there will be no incoherence in thinking, 'Someone might possess states with such-and-such functional role / intentional content while lacking *this* type of state' – where the indexical *this* expresses a concept that is purely recognitional, with no conceptual connections to causal-role concepts or intentional concepts. See Carruthers, 2004a.

to explain why our conscious experiences should seem especially *ineffable* and private, why they should seem to possess intrinsic (non-relational and non-intentional) properties, and so on.

Is it also a *desideratum* of a successful theory, that conscious experiences should be shown to be somehow self-referential in character? While not in the usual catalog of things to be explained, it is arguable that the answer to this question is, 'Yes', in each of two distinct senses. First, it is plausible that the contents of perceptual experience contain an implicit reference to the self (Bermúdez, 1998). Objects are seen as being closer or further away, for example, or as being above or below. Closer to or further from what? Above or below what? The only available answer is: oneself. Equally, when one moves through the world there is a distinctive sort of 'visual flow' as objects approach, loom larger, and then disappear out of the periphery of the visual field. This experience of visual flow is normally apprehended as – that is, has as part of its intentional content – motion through a stationary (or independently moving) environment. Motion of what? Again the only available answer is: oneself.

There is also quite a different sense in which it can be argued that conscious experiences involve a sort of self-reference, however. This is not reference to the self (in the manner sketched above), but rather reference to the very same experience itself. For it seems that conscious experiences, in their distinctive subjectivity, somehow present *themselves* to us, as well as presenting whatever it is that they are experiences *of*. So conscious experiences, besides presenting or referring to items in and properties of the world (or of our own bodies), also present or make reference to themselves. On this sort of view, then, an experience of red, besides having the world-presenting content, *red over there*, will also have the self-referential content, *this is an experience of red over there*.

How can these varying demands on a theory of conscious experience best be met? My own view is a version of higher-order representational account. This is developed and defended at length elsewhere (Carruthers, 2000, 2004a, 2004b); here there is space for just the barest sketch. What constitutes an experience as phenomenally conscious, in my view, is that it possesses a dual representational content: both world (or body) representing and experience representing. And experiences come to possess such a dual content by virtue of their availability to a higher-order thought faculty (which is capable of entertaining higher-

order thoughts about those very experiences), and by virtue of the truth of some or other form of 'consumer semantic' account of intentional content.

There are a number of components that need to be set in place in order for this account to work. First, we need to accept that the intentional content of all perceptual states (whether conscious or unconscious) is non-conceptual, or at least *analog* or fine-grained, in character. Many in recent years have defended the reality of non-conceptual intentional content (Bermúdez, 1995; Tye, 1995, 2000; Kelly, 2001; Luntley, 2003). And even if one feels that these claims may go too far, and that the contents of perception are always to one degree or another *imbued with* concepts, still it needs to be recognized that perceptual experience is always *analog* in relation to any concepts that we can possess (Carruthers, 2000; Kelly, 2001). Our color experiences, for example, have a fineness of grain that far outstrips our capacity to conceptualize, recognize, and remember them. The same holds for movement and shape; and similar things are true in all other sense modalities.

The second necessary ingredient is acceptance of some or other form of consumer semantics. What all kinds of consumer semantics have in common is a commitment to the idea that the intentional content of a state depends in part on what the 'down-stream' consumer systems for that state are apt to do with it or infer from it.[3] (Teleo-semantics is one form of consumer semantics; see Millikan, 1984, 1989; Papineau, 1987, 1993. Functional or inferential role semantics is another; see Loar, 1981; Block, 1986; McGinn, 1989; Peacocke, 1992.) In fact the only kind of semantic theory that *isn't* a form of consumer semantics, is pure input-side, causal co-variance, or 'informational' semantics (Fodor, 1990).

These two main ingredients then need to be put together with what many consider to be a plausible architecture for human cognition, in which perceptual contents are widely 'broadcast' and made available to a variety of down-stream consumer systems for conceptualizing and drawing inferences from those contents (Baars, 1988, 1997). Included

[3] The metaphor comes from conceiving of cognition as a *stream* flowing from input (sensory stimulation) to output (action). Our perceptual systems are 'up-stream', constructing representations as output that are taken as input by (that are consumed by) a variety of 'down-stream' inferential systems, belief-forming systems, planning systems, and so forth. The latter in turn produce representations that are eventually consumed by the motor-control systems.

amongst the latter will be a higher-order thought faculty capable of deploying concepts of experience. And then what we get is the account sketched above. Each perceptual representation with the analog content red_a[4], for example, acquires the higher-order analog content *seems-red*$_a$ or *experience-of-red*$_a$, by virtue of its availability to a higher-order thought system capable of judging immediately and non-inferentially that one is experiencing red.[5]

Such an account can meet all of the main demands made on a theory of conscious experience. First, it can explain how conscious experiences have a subjective dimension of *what it is likeness*. This is their higher-order analog content, in virtue of which they themselves (and not just the objects and properties that their first-order contents are *of*) are presented to us non-conceptually or in analog fashion. Second, the account can explain the distinction between experiences that are phenomenally conscious and those that aren't. This will be the distinction between perceptual states that are, and those that aren't, made available to our higher-order thought faculty, thereby acquiring a higher-order analog content. Third, the account can explain how we can have purely recognitional concepts of our experiences. These will be recognitional concepts whose application conditions are grounded in the higher-order analog content that attaches to those experiences (Carruthers, 2004a). Fourth, the account can explain why our conscious experiences should seem especially ineffable. This is because the fine-grained character of our awareness of those experiences, mediated by their higher-order analog contents, will seem to 'slip through the gaps' of any of our attempts to describe them in conceptual terms. And so on. (For more

[4] Here and throughout I shall use a subscripted 'a' when referring to perceptual contents, to emphasize their fine-grained analog character.

[5] Note that not *any* form of consumer semantics can be endorsed, if this account is to be plausible. Rather, we need claim that it is only the *immediate* further effects of a state are determinants of its content. Otherwise, if distant inferences were determinants of content, we would face the implausible consequence that our perceptual experiences can have the contents, *ripens-in-July*$_a$, *is-Aunt-Anne's-favorite*$_a$, and so forth. It is fortunate, then, that consumer semantics is especially plausible in respect of the *immediate* inferences that consumer systems are apt to derive from a given state. For example, it seems that part of what fixes the content of '&' as *and*, is a disposition to move from 'P&Q' to 'P' and to 'Q' – but not necessarily any more elaborate disposition to derive '~ (P ⊃ ~Q)'. Thus someone could surely mean *P and Q* by 'P&Q', even though they lacked the inferential capacity to deduce from it '~ (P ⊃ ~Q)'.

extensive discussion, see Carruthers, 2000.)

Notice that on this account there is an important respect in which conscious experiences turn out to be self-referential, in addition to the reference to the self that is implicit in their first-order intentional contents. This flows from the dual content that attaches to them. Conscious experiences of red, for example, aren't just targeted on the worldly property (redness) that is represented in analog fashion by their first-order contents. They are also targeted *on themselves*, via their higher-order analog contents of the form, *experience-of-red*$_a$. So we have a vindication of the intuition that conscious experiences don't just present the world (or our own bodies) to us, but also somehow present themselves to us. This 'presenting' is done via their higher-order analog contents, which represent, and replicate in 'seeming fashion', their first-order contents.

2 The demands on theories of conscious thought

If the *desiderata* for theories of conscious experience and conscious thought were the same, then one would expect that people would need to converge on theories of the same general type in respect of each. But this isn't so. While I myself endorse higher-order theories of both conscious experience and conscious thought, for example, such a combination of views is by no means mandatory. In particular, someone might sensibly combine some kind of first-order account of phenomenally conscious experience, with a higher-order account of the conscious status of thought (e.g. Kirk, 1994). This suggests that the demands on explanation, here, are distinct.[6]

If the *desiderata* for theories of conscious experience and theories of conscious thought were the same, indeed, then one would expect that those who endorse first-order representational theories of the former (Kirk, 1994; Dretske, 1995; Tye, 1995, 2000) should also endorse a purely first-order account of the latter. Not only isn't this the case (Dreske and Tye are silent on the nature of conscious thinking; Kirk endorses a higher-order account), but I suspect, moreover, that first-order accounts of conscious thinking

[6] Is it any argument against imposing different *desiderata* on theories of conscious experience and thought, that 'conscious' appears to be univocal when applied to experiences and thoughts? Not at all. For theories of consciousness aren't theories of the *concept* 'conscious'. The concept can very well be univocal while the phenomena picked out by that concept give rise to different explanatory problems.

aren't even defensible. This can be brought out by considering what first-order theorists might say in response to the widespread evidence of *un*conscious perception and *un*conscious thought (Baars, 1988, 1997; Milner and Goodale, 1995).

In the case of conscious experience the main *desideratum*, as we noted, is to explain the properties involved in *phenomenal* consciousness. And it is always then open to a first-order theorist to respond to alleged evidence of non-conscious experience (blindsight, dual-systems theories of vision, and so forth) by insisting that the experiential states in question are actually phenomenally conscious ones, despite not being *access*-conscious. (That is, despite not being available for the subject to know of or report on directly. Tye, 1995, seems to endorse a view of this sort.) There is nothing incoherent in the idea of phenomenally conscious experiences that subjects aren't aware of themselves possessing (even if such a thing is rather hard to believe).

In the case of conscious thinking, however, there would seem to be no independent target of explanation. For in this case there doesn't seem to be any scope for someone to claim that the 'unconscious' thoughts investigated by psychologists are, really, conscious ones, despite being thoughts of which the subjects lack awareness. In the case of conscious thinking *the* phenomenon to be explained is the way that we (seem to have) immediate and non-inferential awareness of (some of) our own thought processes. And this is because thoughts aren't phenomenally conscious *per se*. Our thoughts aren't *like* anything, in the relevant sense, except to the extent that they might be associated with visual or other images or emotional feelings, which will be phenomenally conscious by virtue of their quasi-sensory status.[7]

There is, of course, *a* sense in which it is *like* something to entertain a conscious thought. This is that, depending on what one is thinking about, different aspects of the world thought about will loom into focus. As one's thoughts shift from one topic to

[7] Admittedly, if 'inner speech' can be a kind of thought, as I am inclined to believe, and as we shall discuss briefly in section 6.2, then some thinking *will* have phenomenal properties. These will be the properties, namely, associated with the auditory images that constitute the stream of inner speech. But even in this case it won't be *qua* thoughts that the events in the stream are phenomenally conscious. Phenomenal consciousness will attach to the imaged *sounds* of the sentences in inner speech, not to the contents of those sentences, i.e. not to the thoughts that are thereby entertained.

another, so one's attention shifts from one aspect of the world to another. Siewert (1998) believes that this supports the view that non-imagistic thought is phenomenally conscious. But this is to conflate *worldly* subjectivity with *mental-state* subjectivity (Carruthers, 1998). Of course *the world* is *like* something to any perceiver and to any thinker, whether their states are phenomenally conscious or not. For any experience, and any thought, will involve a partial and partially-subjective 'take' on the objects of experience / thought. What is crucial for phenomenal consciousness, however, is that there should be something that the subject's *own mental states* are *like*, for them. It is the mental states themselves that are subjective in character, that possess properties that are available to introspective recognition, and so on. With this distinction in place, there is no reason to believe that non-imagistic thoughts will be *like* anything.

The only remaining puzzle about conscious thinking, in fact (given that such thinkings aren't necessarily and intrinsically phenomenally conscious) is that we seem to have immediate and non-inferential awareness that we are doing it. So we might as well say that conscious thoughts *are*, then, the thoughts that we can be immediately aware of possessing. Or so, at any rate, I propose to assume in what follows. Our question will be: how is such non-inferential awareness of our own thought processes even so much as *possible*? We will begin on this question shortly, in section 3.

Before we get to that, however, recall the familiar distinction between thoughts as standing states and thoughts as occurrent events (acts of thinking). What is it for beliefs and desires (qua standing states) to be conscious? One proposal, that might seem to flow directly from the assumption we have just made, would be as follows. We might say that standing states are conscious provided that the subject has immediate non-inferential awareness of them. This won't do, however, for a variety of reasons. One (the only one I shall discuss) is that there exist a range of convincing counter-examples, drawn from both Freudian-style psychology and common sense. These are cases where a standing-state belief or desire is the target of seemingly non-inferential higher-order awareness, but without thereby being conscious.

Suppose, for instance, that in a discussion of the merits and demerits of utilitarianism, someone points out to me that I have not only been putting forward utilitarian views, but that I have been speaking of utilitarians as 'we', and have felt

threatened and become angry when utilitarian views as such are maligned. This might strike me with the force of self-discovery. Had I been asked whether I was a utilitarian previously, I would have denied it. I did not *consciously* believe in the truth of utilitarianism. Yet my behavior suggests both that I *believe* utilitarianism to be the correct moral theory, and that I have second-order awareness of this belief (hence the fact that I feel threatened when utilitarian views are attacked).

A better answer to the question of what renders standing-state beliefs and desires conscious would be this: they are conscious just in case they are apt to emerge as conscious occurrent thinkings with the very same first-order content. This is why I didn't consciously believe in utilitarianism, in the example above: I wasn't disposed to think consciously and spontaneously, 'Utilitarianism is true', or something to that effect. This answer also fits neatly with what Gordon (1996) has defended as the 'question / check procedure' for self-ascribing beliefs.[8] If someone asks you whether or not you believe something, what do you do? You surely ask yourself, 'Is it true that P?', and you ascribe the belief to yourself just in case you find yourself inclined to answer, 'Yes, it is the case that P'. In effect, you use your conscious occurrent judgment with the first-order content *P* as the basis on which to ascribe to yourself the standing-state belief that P.

It is plausible that the conscious status of standing-state thoughts should be explained in terms of that of their occurrent counterparts, then. At any rate (again), this is what I propose to assume in what follows. So we will need to focus on what it is for an occurrent act of thinking to be conscious. Here is very natural proposal: a conscious act of thinking is one whose occurrence and content the subject has immediate and non-inferential awareness of (Rosenthal, 1993; Carruthers, 1996).[9] The awareness in question

[8] Note that endorsing this thesis needn't involve any commitment to Gordon's 'simulation theory' of the basis on which we ascribe mental states generally. Endorsing the 'question / check procedure' as the basis on which we have self-awareness of standing-state beliefs is perfectly consistent with some or other version of 'theory-theory' of the overall basis of mental-state ascription.

[9] Note that a major difference between the two authors cited concerns the question whether the higher-order awareness has to be actual, or whether it can be merely dispositional. (There is a perfectly respectable sense in which I can be said to be aware that zebras in the wild don't wear overcoats, of course, or to be aware that ten million and one is larger than ten million, even if I have never explicitly considered and endorsed

surely has to be non-inferential, since otherwise judgments that I attribute to myself as a result of self-interpretation would count as conscious ones. While there is no doubt much that could be said in support of (or against) such a proposal, for present purposes I shall simply assume its correctness, and see what then follows concerning the likely reality of, and the self-referential status of, conscious thinking.

3 How is conscious thinking possible?

Can we describe a possible functional architecture that might serve to realize conscious occurrent judgment, in the sense explained above? What we need is that whenever a judgment of a certain type is being made (e.g. occurring at a particular stage in the stream of processing within the mind's executive or decision-making systems, say), then that judgment is disposed to cause or give rise to the higher-order judgment that just such a judgment is occurring. And such causation needs to be direct, in a way that doesn't count as inferential or self-interpretive.

How might such an architecture be possible? And how might it be realized? Suppose that there is a language of thought, or 'Mentalese'. Then when a sentence in this language, |P|,[10] is entertained at a particular point in processing, we can suppose that the system has been built in such a way that the subject is automatically disposed (if relevant, i.e. depending on what else is going on in the subject's mind) to token the sentence |I am thinking that P|. And provided that the different causal roles distinctive of belief, desire, and so forth are signaled explicitly by operators in the language of thought, then the very same sort of mechanism will also yield non-inferential awareness that I am judging (factively) that P, or that I have an activated desire for P, and so on.

In functional accounts of cognition, beliefs and desires are generally represented by distinct *boxes*. But even if something like this were literally true, it would still have to be the case that token activated beliefs and token activated desires can interact with one another within other systems, such as in practical reasoning. So they would have to be

---

these propositions. I may be said to be aware of these things because I *would* immediately assent to them if I *were to* consider them.) I shall be assuming the latter in what follows.

[10] I shall use line-brackets when referring to sentences in the language of thought / Mentalese, using quote-marks when referring to natural language sentences, and italics when referring to sentence contents.

tagged somehow to indicate which 'box' they derive from. What we might have, then, is the belief that P realized by a Mentalese representation |BEL- P| and the desire for P realized by |DES- P|, where the tags |BEL-| and |DES-| determine their causal roles as beliefs and desires respectively.[11] And then a mechanism can easily be imagined that would go immediately from the former to |BEL- I am entertaining the belief that P| and that would go immediately from the latter to |BEL- I am entertaining the desire that P| – where these would of course mean that I am aware that I am *judging* that P, and that I am aware that I am occurrently *wanting* that P, respectively.

Notice, too, that such an architecture (together with the truth of some version of consumer semantics of the sort appealed to in the explanation of phenomenal consciousness in section 1) might entail that conscious judgments, as well as conscious experiences, are events with a dual intentional content. For the availability of the judgment *P* to a consumer system apt to judge, immediately and non-inferentially, *I am judging that P*, might be sufficient for the initial first-order judgment to acquire a higher-order content. And then one and the same token act of thinking would possess the dual contents *P* and *I am judging that P*.

4 Is conscious thinking actual?

I have outlined an architecture that would vindicate the reality of conscious thinking, while at the same time entailing (given consumer semantics) that conscious thinkings are self-referential. The evidence suggests, however, that the human mind may contain no such architecture as the one just sketched above. For there is now widespread evidence that humans routinely *confabulate* explanations of their own behavior, as has emerged again and again over the last quarter century of social-psychological and neuro-psychological research. (For recent reviews, see Gazzaniga, 1998; Wilson, 2002.) Such data are in tension with the existence of the sort of non-inferential thinking-attribution mechanism

[11] Note that the representation |BEL- P| isn't yet a higher-order one. It isn't a representation *that* the subject believes that P. Rather, it is *constitutive* of the subject believing that P. The tag |BEL-| *causes* other systems to *treat* the representation in the manner constitutive of belief (e.g. by feeding it to inferential systems, or by feeding it to the practical reasoning system to guide action). It doesn't *represent that* the representation in question is a belief.

envisaged above. (Some attempts to render them consistent will be considered in a moment.)

Let me quickly sketch a couple of highlights from this body of research. In one of the classic experiments of Nisbett and Wilson (1977), subjects in a shopping mall were presented with an array of four sets of items (e.g. pairs of socks or panty-hose), and were asked to choose one of them as a free sample. (All four sets of items were actually identical.) Subjects displayed a marked tendency to select the item from the right-hand end of the display. Yet no one mentioned this when they were asked to explain why they had chosen as they did. Rather, subjects produced plainly-confabulated explanations, such as that the item they had chosen was softer, that it appeared to have been better made, or that it had a more attractive color.

As Nisbett and Wilson (1977) point out, what seems to happen in such cases is this. Subjects have a right-hand bias, leading them to spend a longer time attending to the right-most item. Their higher-order thought faculty, noticing and seeking to explain this behavior, proposes an explanation. For example: I am attending more to that item because I believe it to be the softest. And this explanatory higher-order belief is then the source of the subject's verbal report, as well as the subject's choice. But the subject has no access to the process of interpretative thinking that generated their higher-order belief; and that belief itself is without any foundation in the first-order facts – it certainly isn't produced by the sort of non-inferential ascent-mechanism envisaged in section 3.

The second example is reported in Gazzaniga (1998), concerning one of his 'split brain' patients. When the instruction, 'Walk!', was flashed up in the patient's left visual field (accessible to his right hemisphere, which had some capacity to comprehend simple forms of language, but no productive abilities), the patient got up and started to walk across the room. When asked what he was doing, he (his left hemisphere, which controls speech) replied, 'I want to get a coke from the fridge'. This answer was plainly confabulated, generated by his higher-order thought faculty (which independent evidence suggests is located largely in the left hemisphere) in order to explain his own behavior. But the answer came to him with all of the obviousness and apparent indubitability that attaches to any of our ascriptions of occurrent thoughts to ourselves.

The thoughts that actually produced the subject's behavior, in this example, were

presumably |DES- I comply with the experimenter's instruction| and |BEL- To comply with the instruction to walk, I must walk|. Whereas the higher-order thought faculty, being aware of the subject's own behavior and seeking to explain it, came up with the explanation |BEL- I am walking because I want to get a coke from the fridge| (perhaps noticing that the fridge lay in the direction that he was walking). And the higher-order attribution of desire, here, was plainly an inference-produced product of self-interpretation, not resulting from the operations of some sort of ascent-mechanism.

This and similar data lead Gazzaniga (1998) to propose that the left hemisphere of the brain houses an 'interpreter' (a higher-order thought faculty), which has access to perceptual input, but not to the occurrent conceptual thoughts and decision-making processes occurring elsewhere in the brain. The interpreter is continually weaving an explanatory story for the agent's own actions. These stories may often be true ones, in familiar-enough cases and in cases where the interpreter does its job well. But they are still a product of interpretation, and not the result of any sort of non-inferential access to the subject's own thought processes. And in unusual or unexpected circumstances the subject may end up with stories that are confabulated (i.e. false).

If any such account is true, then a plausible abductive inference – in this case an application of Occam's razor – suggests that the human mind does *not* have the sort non-inferential semantic-ascent architecture that we sketched in section 3. And it appears to follow, too (if these cases can be taken as representative) that there is no such thing as conscious thinking.

Can such a conclusion be ruled out of court immediately, rejected on the grounds that we can be *certain* that there is such a thing as conscious thinking? No, it can't. For we are assuming that conscious thinking requires non-inferential awareness of our own thought processes. But all we can be certain of – the most that introspection can deliver – is that we are sometimes aware of our own thought processes without engaging in any *conscious* inference. We can't be certain that our awareness of our own thought processes isn't grounded in an *un*conscious inference. And if Gazzaniga is right, it always is.

It is worth noting that Gazzaniga's proposal is consistent with, and to some degree receives independent support from, an overall architecture for cognition that has been receiving increasing support in recent decades (Baars, 1997; Carruthers, 2000 ch.11, 2002).

On this account the various sensory systems produce integrated analog representations of the environment (and body), which are then widely broadcast and made available to a range of down-stream conceptual systems (for higher-order thought, for folk mechanics, for folk biology, and so on). These latter systems have quite limited access to one another, however. (They are to some degree 'encapsulated'.) And neither do they have access to what takes place even further down-stream, within practical reasoning. So on this model, although the higher-order thought faculty would have access to perceptual and proprioceptive input (and hence to whatever the agent is physically doing), it won't have any direct access to the thought processes that cause our actions. I shall return to discuss this model at greater length in section 5.

One way in which it might be argued that the confabulation data are consistent with an architecture of the kind sketched in section 3, however, would be this. Perhaps the confabulated judgments are made too long after the event to be reliable, or for the semantic-ascent architecture envisaged in section 3 to operate. It is plausible enough that the decay-time for any given occurrent thought should be pretty brief. So if the token Mentalese sentence |P| doesn't give rise to |I am thinking that P| almost immediately, the subject will have no option but to self-interpret; which might lead, in the right circumstances, to confabulation. This reply doesn't really work, however. For a subject can be asked for an explanation immediately after she makes a choice (in the Nisbett and Wilson example), or while he is getting up out of his chair (in the Gazzaniga example). And the window for unrehearsed items to remain in working memory isn't generally reckoned to be *this* brief.

A related objection would be this. There are a range of experimental demonstrations that so-called 'think aloud protocols' – in which subjects verbalize their thinking out loud *while* reasoning to the solution of some problem – are really quite reliable in providing us with a window on the underlying sequences of thought in question (Ericsson and Simon, 1993). And how can this be possible unless those subjects have reliable (non-confabulated) awareness of the thoughts that they verbalize? But in fact, linguistic *expression* of a thought need not imply that the subject has higher-order awareness that they are entertaining that thought. And indeed, one of the central findings in this area is that subjects need to be induced *not* to report *on* their thoughts when they have

them, since this is demonstrably *un*reliable (Ericsson and Simon, 1993).

Notice that the production sub-system of the language faculty will need to be situated down-stream of the various belief-forming and decision-making reasoning processes that figure in cognition, so that the results of those processes should be expressible in speech (Carruthers, 2002). And although one of these systems that feeds input to the language faculty will be the higher-order thought faculty, there is no reason to assume that the language faculty can *only* receive higher-order thoughts as input. On the contrary, many of our first-order thoughts should be *directly* expressible in speech. This is sufficient to explain the Ericsson and Simon data. But then unless the linguistic expressions of thought are somehow constitutive of the thoughts being articulated, our awareness of what we are thinking will be derivative from our awareness of the sentences in which those thoughts are expressed – and it looks as if this won't, then, have the kind of immediacy required for those thoughts to count as conscious ones. (We return to some of these points in section 6.)

Another way in which someone might try to argue that the confabulation data are consistent with the required sort of semantic-ascent architecture would be this. Perhaps in the confabulation cases the thoughts in question don't occur in the right sub-system. Perhaps there are two distinct sub-systems in the mind in which thinking occurs, and which can generate behavior. But perhaps only one of these has the kind of direct access to the higher-order thought faculty that we envisaged earlier. So the thoughts in this sub-system would be conscious; whereas the confabulation behaviors are produced by the other sub-system, whose contents *aren't* conscious. However, it is hard to see any plausible way of drawing the sub-divisions here, that wouldn't simply be *ad hoc*.[12] For the confabulation examples seem pretty much like paradigmatic cases of (non-natural-language-based) judgment.

[12] One suggestion – which definitely isn't *ad hoc*, since it is supported by multiple lines of evidence – would be to claim that there are dual systems for thinking and reasoning, one of which is fast, implicit, and unconscious, and the other of which is slow, explicit, and conscious (Evans and Over, 1996; Stanovich, 1999). However if, as some have argued, the explicit system implicates natural language sentences (Evans and Over, 1996; Frankish, 2004), then it won't exemplify the sort of Mentalese-based semantic-ascent architecture that is under discussion here. This point will be further explored in section 6.

It would appear, then (if the problem of sub-divisions can't be overcome), that the confabulation evidence will show that we don't have the kind on non-inferential access to our own acts of thinking for those acts to count as conscious ones. And nor, either, will our acts of thinking have the right sort of self-referential content. For if the thought *P* isn't available to a higher-order thought faculty that is disposed to judge immediately that I am thinking that P, then the thought *P* won't at the same time bear the higher-order self-referential content *I am thinking that P*.

5 An alternative model of the higher-order thought faculty

Let us assume that the problem of sub-divisions can be overcome, however. Assume that there is some non-arbitrary way of distinguishing between those reasoning systems whose processes are directly available to higher-order thought, and those that aren't. Then what we have on the table is an alternative model of the way in which a higher-order thought faculty could be embedded into the overall architecture of the mind, to be contrasted with the model deriving from Baars (1997) sketched above. According to the latter, the higher-order thought faculty has direct access *only* to those of our occurrent states that are perceptual, necessary to construct explanations and predictions of people's behavior.[13] Call this the 'mind-reading model'. According to the alternative now being suggested, the higher-order thought faculty *also* has direct access to some of the other reasoning processes taking place down-stream of perception, especially some of the processes that occur within *practical* reasoning. Call this the 'self-monitoring model'.

These two models correspond to two different accounts of what higher-order thought is *for*. According to the mind-reading model, higher-order thoughts are for interpreting and predicting behavior. The mind-reading faculty evolved in highly social creatures (such as our great-ape ancestors manifestly were) for purposes of manipulation, co-operation, and communication. This is the standard explanation that cognitive scientists offer of the evolution of our capacity for higher-order thought (e.g., Byrne and Whiten, 1988, 1998). And on this account, the application of higher-order thoughts to ourselves,

[13] The higher-order thought faculty would also need access to (activations of) standing-state beliefs, of course, such as beliefs about the target-subject's long-term goals or idiosyncratic beliefs. But this wouldn't require it to have access to the processes within the agent that generate beliefs and decisions.

and the dual-analog-content that consequently comes to attach to our perceptual states, is an evolutionary spin-off.

The self-monitoring model, in contrast, will claim that higher-order thought is *also* for monitoring our own processes of thinking and reasoning – enabling us to trouble-shoot in cases of difficulty or breakdown, and enabling us to reflect on and improve those processes themselves. (It could be claimed *either* that this is the *basic* function of our higher-order thought faculty, and that a capacity to predict and explain behavior came later; *or* that the mind-reading and self-monitoring functions of the faculty co-evolved.) Some cognitive scientists have begun to explore just such an hypothesis (e.g., Smith *et al*., 2003).

There are some strong *prima facie* reasons for preferring the mind-reading model to the self-monitoring model, however. The most important is that the former appeals to what is, uncontroversially, a highly-developed cognitive competence, whereas the latter doesn't. Everyone agrees that human beings are quite remarkably good at predicting and explaining the behavior of themselves and others through the attribution of mental states. And everyone agrees that this capacity forms part of our natural endowment, emerging in any normally developing member of the species. In contrast, it is *very* controversial to claim that humans have any natural competence in correcting and improving processes of reasoning. On the contrary, both common sense and cognitive science are agreed that naïve subjects are extremely *poor* at spotting errors in reasoning, and at seeing how to improve their own reasoning.[14]

These issues are too large to pursue in any detail here. And to the extent that they remain unresolved, the self-monitoring model (combined with the semantic-ascent architecture envisaged in section 3) holds out the hope that we may yet be shown to engage

[14] People *can* monitor their own reasoning, for course, even if they aren't very good at improving it (although they can get better) – especially when their reasoning is verbally expressed. But this lends no support to the version of self-monitoring model under discussion here. For the best account of this capacity is that it is *realized in* cycles of operation of other systems (including language and mind-reading), and that it is – like Dennett's 1991 *Joycean machine* – heavily influenced by cultural learning (Carruthers, 2002; Frankish, 2004). By learning to verbalize our own thoughts we can learn to monitor and improve upon our own patterns of reasoning. But only if our verbalizations are constitutive of (a kind of) thinking will our access to our own thoughts count as immediate and non-inferential. (See the discussion in section 6.)

in conscious thinking independently of the use of sensory images. In what follows, however, I shall assume that the mind-reading model of our higher-order thought abilities is the correct one. This is partly because interesting questions then arise, concerning the extent to which sensory images could nevertheless underpin a kind of conscious propositional thinking. And it is partly because it is worth exploring what would follow if the self-monitoring model turns out to be false, since it may well turn out to *be* false. And in philosophy, of course, the conditional questions are often the most interesting and important ones.[15]

6 Does inner speech make thinking conscious?

So far, then, the evidence looks as if it might point to the conclusion that there is strictly speaking no such thing as conscious thinking (at least, to the extent that thinking isn't expressed in natural language or other imagery). And some cognitive scientists have concluded just this (even if not in exactly these words; see Gopnik, 1993). But what of 'inner speech', however? Might this give us the kind of immediate awareness of our own thought processes to constitute some of the latter as conscious?

Our discussion of these questions now needs to proceed in two parts, corresponding to the contrast that I have drawn elsewhere between 'communicative' and 'cognitive' conceptions of the role of natural language in cognition (Carruthers, 1996, 2002). According to the communicative conception of language, the only real function of language is communication (whether with another or with oneself). Natural language sentences *express* thought, but aren't *constitutive of* thought. According to the cognitive conception of language, on the other hand, at least some of our thinking takes place in natural language. So on this view, natural language sentences are, at least sometimes, (partly) constitutive of acts of thinking. Let us take these possibilities in turn, the former in section 6.1 and the latter in section 6.2.

6.1 Inner speech as expressive of thought

Consider first, then, the traditional view that inner speech is *expressive* of thought, rather than directly (and partly) *constitutive of* it. On this account, thinking itself might be

[15] As one of the characters in the Walt Disney movie *Hercules* remarks to another, '*If* is good!'

conducted in some sort of Mentalese. (Let us assume so.) But some of these Mentalese representations can be used to generate a representation of a natural language sentence in auditory imagination, creating the phenomenon of inner speech. Might this be sufficient to give us the sort of non-inferential awareness of the underlying thoughts that is required for the latter to count as conscious ones?

Suppose that the contents of the Mentalese acts of thinking and the contents of the natural language sentences generated from them line up neatly one-for-one. Then thinking something carried by the Mentalese representation |BEL- P| will cause a suitable (indicative-mood) natural language sentence 'P' to be imaged, where the contents of |P| and 'P' are the same. But we might suppose that the imaged sentence 'P' comes with its semantic properties somehow attached – for after all, when we form an image of a sentence, we don't just hear imaginary *sounds*, we also (as it were) hear *the meaning*, just as we do in normal speech comprehension.

Then suppose that I am disposed to move from the imaged sentence 'P' to the higher-order representation |I am thinking that P|, in which the content of the representation 'P' is extracted and re-used within the content of the that-clause. It will then turn out that it is pretty much guaranteed that such self-attributions will be reliable. Moreover, the immediacy of the causal pathway involved could be sufficient for the higher-order item of awareness in question to count as non-inferentially produced; in which case the first-order thought that *P* could count as conscious. By the same token, too, that thought might qualify as having a dual content, making conscious thinking self-referential in something like the way that conscious experiences are (on my account).

There are two significant problems with this neat picture, however. The first is that, even if self-attributions of thought *contents* resulting from the use of inner speech are immediate (non-self-interpretative and non-inferential), self-attributions of thought *modes* (such as judging and wanting) surely aren't. This is because natural language sentences don't wear their modes on their face.

An utterance of the indicative sentence, 'The door is open', can in suitable circumstances express the *belief* that the door is open, or ask a *question* as to whether or not the door is open, or issue a *command* to close the door, or merely express the *supposition* that the door is open for purposes of further inference, and so on. So whether

or not an indicative-mood sentence in inner speech, ‘P’, is expressive of the subject’s *judgment* (i.e. occurrent belief) that P, simply cannot be recovered from the sentence alone. It is always going to be a matter of self-interpretation to attribute to oneself a given judgment, on this sort of account. And that seems sufficient to disqualify such judgments from counting as conscious ones.

It might be replied that in spoken language, *mode* is often indicated by tone of voice; and this can be amongst the contents of the auditory images that figure in inner speech. So the basis for my knowledge that I am *judging* that P when I token the natural language sentence ‘P’ in auditory imagination, is the imagined tone of voice in which that sentence is ‘heard’. This reply won’t wash, however, for two distinct reasons. The first is that although the mode in which a sentence is meant *can* be marked by intonation, in needn’t be – someone’s delivery can be entirely neutral, or ‘flat’. So this couldn’t be a quite general solution to our problem. But the second, and more fundamental, reason is that tone of voice must any case be *interpreted* to yield the intended mode. If someone says, ‘The door is open’, in a sharp, angry-sounding voice, for example, it requires interpretation to tell whether they are expressing a *belief* about something that they disapprove of, or are issuing a *command* to close the door. Telling which it is might require knowledge of our respective power / authority relations, among other things.

The second problem for the simple account sketched above is that natural-language sentence-contents and the contents of the Mentalese sentences used to generate them will rarely line up one-for-one. Language routinely makes use of contextual factors in expressing meaning. The sentence, ‘The door is open’, leaves it to the context to determine which door is *the* door; it also leaves it to the context to determine the appropriate standard of openness (unlocked? open just a crack? wide open?); and so on. In contrast, the corresponding sentence of Mentalese must render such facts determinate. So again, one can’t recover the underlying Mentalese thought from the natural language sentence alone.

It might be argued that these problems can be overcome, however, if self-generated sentences (in inner-speech) can somehow carry with them the elements necessary for their interpretation. For then provided that those same meaning-determining connections are also inherited by the higher-order Mentalese sentence that replicates the content of the first-order one within a that-clause, we may still have the sort of immediacy needed for

conscious thinking.

Perhaps it works like this. The underlying assertoric thought with the content *P* is carried by the Mentalese expression |BEL- P|. This is then used to generate a natural language sentence 'Q' in auditory imagination. But that sentence comes with the connections to |BEL- P| already attached. The imaged sentence 'Q', by virtue of being 'experienced', is a state of the right sort to be received as input by the mind-reading faculty, which can deploy the concept of occurrent belief. The mind-reading faculty detaches the Mentalese sentence |BEL- P| from the natural language sentence received as input, and forms from it the Mentalese belief |BEL- I am thinking that P|, in which the Mentalese sentence |P| is re-used with the same content as the original. And the result might then count as non-inferential awareness of my own act of thinking.

I suspect that there may be a good many problems with this attempted vindication of the reality of conscious thinking. Let me focus on one. It is quite widely accepted that the language faculty is divided into two distinct sub-systems, one for production and one for comprehension (with perhaps each of these drawing off a single language-specific data-base; Chomsky, 1995). It will be the work of the production sub-system to create the natural language sentence 'Q' from the Mentalese representation |BEL- P|. But in order for that sentence to be displayed in auditory imagination and received by the mind-reading faculty, it needs to be passed across to be received by the *comprehension* sub-system. And there is good reason to think that the connections with the underlying thought, expressed by |BEL- P|, will thereby be severed.

One reason for this is that the comprehension sub-system simply isn't *built* to receive Mentalese representations as input. Its job is rather to receive natural language sentences as input and to construct interpretations of them, perhaps in co-operation with other systems. Another reason is that 'inner speech' may well exploit the feed-back loops within the overall language faculty that are used during normal speech production for phonological and semantic monitoring and repair (Levelt, 1989). In normal speech production, the sentence 'Q', generated from the Mentalese message-to-be-communicated |BEL- P|, is passed to the consumer sub-system to check that the intended utterance will indeed convey the intended message. This can only work if the consumer system doesn't *already* have access to the message |BEL- P|.

6.2 Inner speech as (partly) constitutive of thought

It appears, then, that if language is but a means of *expressing* thought, then there may be no such thing as conscious thinking. For although we come to be aware of our thoughts by consuming their internal expressions, in inner speech, the route from thought, to speech, to awareness of thought is too indirect and interpretative to constitute the thoughts in question as conscious ones. Everything may look a bit different if we switch to a version of the cognitive conception of language, however (Carruthers, 1996, 2002), according to which inner speech *is*, or is somehow constitutive of, thinking. To be plausible, such a view should only claim that representations of natural language sentences in inner speech are *partly* constitutive of thinking. (This is because of the problems of indeterminacy attaching to natural language sentences, *inter alia*, discussed in section 6.1.)

Within the framework provided by a cognitive conception of language, an account can be given of how we have non-inferential knowledge of the *contents* of our (conscious) thoughts. The sentence 'Q', generated by the production sub-system, is tokened in inner speech and consumed by the comprehension sub-system. The result will be a representation of an interpreted sentence, carrying with it the connections to the Mentalese expressions, underlying data structures, perceptual experiences, or whatever else is necessary to make the meaning of 'Q' determinate. In the simplest case, if the interpretation process is a reliable one, then the meaning that gets attached to 'Q' might be the same as the content of the Mentalese sentence |P| that initiated the production of 'Q'. But this doesn't really matter in the present context. And it might not always happen. So let us work with an example in which it doesn't: let us imagine that the process of interpreting 'Q' attaches to it the Mentalese sentence |R|.

Now by hypothesis (if some version of the cognitive conception of language is correct) the pairing of <'Q', |R|> has further consequences in cognition; and not just *any* consequences, but those that are distinctive of thinking. One way in which this might be the case is if the representation |R| is one that can *only* be formed via the construction of an appropriate natural language sentence, as 'module-integration' accounts of the role of natural language in cognition suggest (Hermer-Vazquez *et al*., 1999; Carruthers, 2002). Another way in which it might be true is if it is only by virtue of articulating the sentence

‘Q’ in auditory imagination, and hence making its content available to the various inference-systems that exist down-stream of perception and consume its products, that the subject comes to believe |R| for the first time. The process of articulating ‘Q’ leads to |R| being evaluated and accepted, in a way that would not have happened otherwise.[16]

Now amongst the consumer-systems to which <‘Q’, |R|> is made available by the language comprehension sub-system will be the mind-reading faculty. Suppose that the latter is immediately disposed, whenever it receives such a pairing, to form the belief |I am thinking that R|. Then the result will be non-inferential awareness of what I am thinking. We can regard the immediacy and reliability of the connection between the higher-order thought and the thought thereby attributed as being sufficient both to render the act of thinking that R conscious, and to mean that the sentence ‘Q’ has both the first-order content *that R* and the higher-order content *I am thinking that R*. So now we have a single event (a token representation of the natural language sentence ‘Q’ in inner speech) that has both a first-order and a higher-order content, similar to the case of experience.

Note that this ‘immediacy’ needn’t be at all undermined by the fact that the comprehension process that generates an interpretation for ‘Q’ is an inferential and interpretative one. For it is the product, rather than the initial cause, of the interpretative process that gets self-attributed. And this can be attributed to oneself *without* further interpretation or inference. According to the hypothesis that we are considering (the cognitive conception of language), the sentence ‘Q’ displayed (and interpreted) in inner speech is *itself* a thought, or is rather partly *constitutive of* a thought, given its causal role in the overall architecture of cognition. And it is *this* thought (the thought expressed by |R|) that gets reliably and non-inferentially attributed.

It would appear, therefore, that if the cognitive conception of language is correct, then we have a vindication of the reality of conscious thinking. For we can have immediate and non-inferential awareness of the contents of those acts of thinking that occur in inner speech, on this account. However, the point that awareness of attitude (as opposed to

[16] This might happen if the subject *avows* ‘Q’, for example – where this means that they *commit themselves* to thinking and reasoning in future as if ‘Q’ were true (Frankish, 2004). If the subject thereafter remembers and executes this commitment, the effect will be that the underlying representation |R| will become the functional equivalent of |BEL- R|.

awareness of content) must always be inferential / interpretative remains in force. Even if the tokening of some natural language sentence ‘Q’ in auditory imagination is sometimes constitutive of thinking, still the fact that the entertaining of that sentence is an assertoric judgment, or a wondering-whether, or an act of supposition, or whatever, will be a matter of its larger causal role *beyond* the point at which interpretation occurs. (It will be a matter of the further causal role of |R|, indeed.) And that role just can’t be read off from the sentence itself. It will have to be a matter of further self-interpretation.

The upshot is that, while there might be such a thing as conscious (and self-referring) *thinking*, there might be no such thing as conscious assertoric *judging*, conscious (propositional) *wanting*, conscious *supposing*, and so forth. Put differently: although there are conscious episodic propositional *contents*, there might be no conscious episodic propositional *attitudes*.

What of the self-referential character of conscious thinking, on this conception? In what sense is it vindicated? As I presented the view above, I tacitly assumed that the higher-order thought generated from the sentence / thought pair <‘Q’, |R|> would be |BEL- I am thinking that R|. That is, I assumed that the sort of self-reference here would be a reference *to the self*. But perhaps this was unwarranted. One can just as well move directly from, <‘Q’, |R|> to |BEL- *That* is an act of thinking that R|. This suggests that (assuming the truth of some form of consumer semantics) the sentence ‘Q’ might have the dual contents *R* and *that is a thinking that R*, where the pronoun refers to the sentence ‘Q’ in question. In that case, conscious thoughts may be self-referential in exactly the same sort of way that conscious experiences are (as discussed in section 1).

7 Conclusion

I have sketched an account of phenomenally conscious experience according to which such experiences always have dual (and self-referential) analog contents. I have argued that the constraints placed on a theory of conscious thinking are different from those placed on a theory of conscious experience, since conscious thoughts aren’t necessarily and intrinsically *phenomenal* in character. I have sketched some reasons for thinking that there might be no such thing as conscious thinking, if natural language plays no direct role in our thoughts, since all self-attributions might then be inferential / self-interpretative ones. And

I have argued that, if language *does* play such a role, then the *contents* of our thoughts might be conscious (and self-referential) even if the *attitudes* that we take to them are not.[17]

References

Baars, B. 1988. *A Cognitive Theory of Consciousness*. Cambridge University Press.

Baars, B. 1997. *In the Theatre of Consciousness: The workspace of the mind*. Oxford University Press.

Bermúdez, J. 1995. Non-conceptual content. *Mind and Language*, 10, 333-369.

Bermúdez, J. 1998. *The Paradox of Self-Consciousness*. MIT Press.

Block, N. 1986. Advertisement for a semantics for psychology. *Midwest Studies in Philosophy*, 10, 615-678.

Byrne, R. and Whiten, A. (eds.) 1988. *Machiavellian Intelligence*. Oxford University Press.

Byrne, R. and Whiten, A. (eds.) 1998. *Machiavellian Intelligence II: Evaluations and extensions*. Cambridge University Press.

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

Carruthers, P. 1998. Natural theories of consciousness. *European Journal of Philosophy*, 6, 203-222.

Carruthers, P. 2000. *Phenomenal Consciousness: a naturalistic theory*. Cambridge University Press.

Carruthers, P. 2002. The cognitive functions of language. & Author's response: Modularity, language, and the flexibility of thought. *Behavioral and Brain Sciences*, 25, 657-719.

Carruthers, P. 2004a. Phenomenal concepts and higher-order experiences. *Philosophy and Phenomenological Research*, 68, 316-336.

Carruthers, P. 2004b. HOP over FOR, HOT theory. In R. Gennaro (ed.), *Higher Order Theories of Consciousness*. Philadelphia: John Benjamins, 115-135.

[17] I am grateful to Georges Rey for a series of conversations that prompted me to begin writing about the main topic of this paper. And I am grateful to Keith Frankish, Uriah Kriegel, and Georges Rey for insightful sets of comments on earlier drafts.

Chomsky, N. 1995. *The Minimalist Program*. MIT Press.

Dennett, D. 1991. *Consciousness Explained*. Penguin Press.

Dretske, F. 1995. *Naturalizing the Mind*. MIT Press.

Ericsson, K. and Simon, H. 1993. *Protocol Analysis: Verbal reports as data*. (Revised edition.) MIT Press.

Evans, J. and Over, D. 1996. *Rationality and Reasoning*. Psychology Press.

Fodor, J. 1990. *A Theory of Content and Other Essays*. MIT Press.

Frankish, K. 2004. *Mind and Supermind*. Cambridge University Press.

Gazzaniga, M. 1998. *The Mind's Past*. California University Press.

Gopnik, A. 1993. How we know our minds: the illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences*, 16, 1-14.

Gordon, R. 1986. 'Radical' simulationism. In P. Carruthers and P. Smith (eds.), *Theories of Theories of Mind*, Cambridge University Press.

Hermer-Vazquez, L., Spelke, E., and Katsnelson, A. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*, 39, 3-36.

Jacob, P. and Jeannerod, M. 2003. *Ways of Seeing*. Oxford University Press.

Kelly, S. 2001. Demonstrative concepts and experience. *Philosophical Review*, 110, 397-420.

Kirk, R. 1994. *Raw Feeling*. Oxford University Press.

Levelt, W. 1989. *Speaking: from intention to articulation*. MIT Press.

Loar, B. 1981. *Mind and Meaning*. Cambridge University Press.

Luntley, M. 2003. Non-conceptual content and the sound of music. *Mind and Language*, 18, 402-426.

McGinn, C. 1989. *Mental Content.* Blackwell.

Millikan, R. 1984. *Language, Thought, and Other Biological Categories.* MIT Press.

Millikan, R. 1989. Biosemantics. *Journal of Philosophy*, 86, 281-297.

Milner, D. and Goodale, M. 1995. *The Visual Brain in Action*. Oxford University Press.

Nisbett, R. and Wilson, T. 1977. Telling more than we can know. *Psychological Review*, 84, 231-295.

Papineau, D. 1987. *Reality and Representation.* Blackwell.

Papineau, D. 1993. *Philosophical Naturalism.* Blackwell.

Peacocke, C. 1992. *A Study of Concepts.* MIT Press.

Rosenthal, D. 1993. Thinking that one thinks. In M. Davies and G. Humphreys (eds.), *Consciousness*, Blackwell.

Siewert, C. 1998. *The Significance of Consciousness*. Princeton University Press.

Smith, J., Shields, W. and Washburn, D. 2003. The comparative psychology of uncertainty monitoring and meta-cognition. *Behavioral and Brain Sciences*, 26.

Stanovich, K. 1999. *Who is Rational? Studies of individual differences in reasoning*. Laurence Erlbaum.

Tye, M. 1995. *Ten Problems of Consciousness*. MIT Press.

Tye, M. 2000. *Consciousness, Color and Content*. MIT Press.

Weiskrantz, L. 1997. *Consciousness Lost and Found*. Oxford University Press.

Wilson, T. 2002. *Strangers to Ourselves*. Harvard University Press.