Peter Carruthers

The Illusion of Conscious Thought

Abstract: *This paper argues that episodic thoughts (judgments, decisions, and so forth) are always unconscious. Whether consciousness is understood in terms of global broadcasting/widespread accessibility or in terms of non-interpretive higher-order awareness, the conclusion is the same: there is no such thing as conscious thought. Arguments for this conclusion are reviewed. The challenge of explaining why we should all be under the illusion that our thoughts are often conscious is then taken up.*

Keywords: attention; confabulation; consciousness; self-knowledge; thought; working memory.

1. Introduction

For present purposes, *thought* will be understood to encompass all and only propositional-attitude events that are both episodic (as opposed to persisting) and amodal in nature (having a non-sensory format). Thoughts thus include events of wondering whether something is the case, judging something to be the case, recalling that something is the case, deciding to do something, actively intending to do something, adopting something as a goal, and so forth. But thoughts, as herein understood, do not include perceptual events of hearing or seeing that something is the case, feelings of wanting or liking something, nor events of episodic remembering, which are always to some degree sensory/imagistic in character. Nor do they include episodes of inner speech, which may *encode* or *express* thoughts in imagistic format, but which are not themselves attitude events of the relevant kinds. I propose to argue, not only that thoughts *can* be unconscious, but that

Correspondence:
Email: pcarruth@umd.edu

they are *always* unconscious. At the same time, I will explain how we come to be under the illusion that many of our thoughts are conscious ones.

Almost everyone *believes* that thoughts can be conscious, no matter whether consciousness is defined in terms of global accessibility or in terms of non-interpretive higher-order awareness. It seems obvious that our thoughts sometimes occur in a way that makes them widely accessible to other systems, for forming memories, for issuing in positive or negative affect, for guiding decision making, and for verbal report. This would make them first-order access-conscious. But it also seems obvious that those same thoughts are available in a way that enables us to know of their occurrence without requiring self-interpretation, of the sort that makes us aware of the thoughts of other people. This would make them higher-order access-conscious.[1]

I have argued elsewhere that both views are mistaken. In Carruthers (2011) I argue against the second of these accounts, showing that our knowledge of our own thoughts is always interpretive, grounded in awareness of both our own overt behaviour and covert sensory cues of various sorts (visual imagery, inner speech, and so on). The main focus of Carruthers (2015a), in contrast, is to argue that the only mental states that can be globally broadcast (and hence become first-order access-conscious) are those that have a sensory grounding of some kind (including visual and auditory imagery as well as inner speech). So on neither account of consciousness are thoughts themselves ever conscious.

---

1 Some philosophers who endorse a higher-order account of consciousness allow that the relationship between the conscious state and one's knowledge of it can be inferential (Rosenthal, 2005). But what this generally means is that there is some computational process that leads from the state itself to one's higher-order access to it (much as there is a computational process that leads from patterns of light stimulating the retina to representations of a 3-D world, which can also be characterized as 'inferential'). It is not envisaged that the process is *interpretive* in the way that one's knowledge of other people's mental states is, drawing on observations of behaviour together with physical and social circumstances. Other philosophers think that the relationship between thought and awareness of thought (even if interpretive) is that the latter is partly *constitutive* of the former (Schwitzgebel, 2002; 2011). On this view, beliefs, in particular, are said to be clusters of dispositions, included among which are dispositions to have self-knowledge. For a critique of such views, see Carruthers (2013a), which endorses a strongly representationalist account of belief. I will assume, here, that thoughts are not dispositions but structured entities, whose causal roles are sensitive to their structural properties.

In what follows I briefly review both sets of arguments against the existence of conscious thought. In Section 2 I argue that all knowledge of our own occurrent thoughts is interpretive in character, similar to the access that we have to the thoughts of other people. In Section 3 I argue that global broadcasting depends upon attentional signals directed at mid-level sensory areas of the brain, implying that only events with a sensory-based format can be access-conscious.[2] Then in Section 4 I take up the question of how we come to be under the illusion of conscious thought. How is it that nearly everyone believes that there are conscious thoughts if really there aren't? Providing a satisfactory answer to this question is the main goal of the paper.

It should be noted, however, that there are alternative theoretical accounts of consciousness besides the two that will form our focus here. In addition to global broadcasting accounts (Baars, 1988; 2002; 2003; Dehaene *et al*., 2006; Dehaene, 2014) and higher-order access theories (Carruthers, 2000; Rosenthal, 2005; Graziano, 2013), there is Tononi's integrated information account of consciousness, for example (Tononi, 2008; Tononi *et al.*, 2016). I shall ignore the latter for present purposes. In part this is because it is only a theory of *phenomenal* consciousness, and makes no commitments concerning the relevant accessibility relation for conscious mental events (indeed, some might see this as a fatal weakness, since it seems to allow for multiple forms of highly integrated informational state that aren't accessible to their subjects). In fact, my focus here is *only* on so-called 'access consciousness'. Our question is whether thoughts are ever access-conscious, in either a first-order or a higher-order sense. If they aren't, then most people would agree that they can't be phenomenally conscious either. But even if they are, it is much more controversial to claim that thoughts can also be *phenomenally* conscious, or intrinsically *like* something to undergo. I shall say nothing about that here. (For a critique, see Carruthers and Veillet, 2011.)

---

2 In this I follow Prinz (2012), albeit using additional arguments. Mid-level sensory areas in vision would include extrastriate regions V2, V3, V4, and MT (but not primary cortical projection area V1), which process visual stimuli for contrast, shape, colour, and movement. Processing that underlies category recognition takes place in high-level visual areas, which in the case of vision would include the lateral and ventromedial temporal cortex.

2. Interpretive Self-Knowledge

How do we know what we are currently thinking? Intuition has it that such knowledge is (often) immediate. One merely has to introspect in order to know that one has just decided to do something, or to know what one currently believes when asked a question. Importantly, our knowledge of our own thoughts is believed by most philosophers to differ in *kind* from our knowledge of the thoughts of other people. One knows what someone else is thinking by observing and drawing inferences from their circumstances and behaviour (including their speech behaviour). All such knowledge is believed to be interpretive, using one's 'theory of mind' or 'mind-reading' skills to infer the mental states that lie behind the other person's observable behaviour. These inferences needn't be conscious ones, of course. Indeed, as a matter of phenomenology one often just seems to *intuit* or *see* (or *hear*, in the case of speech) what someone is thinking in a particular context. But most would maintain that such intuitions are nevertheless grounded in one's knowledge of the likely causes of the behaviour one observes.

While most philosophers and psychologists think that one's knowledge of the thoughts of others is at least tacitly interpretive, drawing on background knowledge provided by some sort of folk psychology, not everyone agrees. Some think that knowledge of other minds can be more directly perceptual (at least in simple cases), perhaps responding to behavioural and environmental *affordances* of a social-interactive sort (Gallagher, 2001; Hutto, 2004; Noë, 2004). Such views seem to me ill-motivated. For on closer examination they fail to offer a plausible route through which perceptual knowledge of other minds can be achieved (Spaulding, 2016). Moreover, one can in any case explain the largely intuitive nature of much of our knowledge of other minds within a classical knowledge-based framework. Indeed, it is possible to endorse such a framework while claiming that our awareness of other people's mental states is genuinely perceptual in character (Carruthers, 2015b). In addition, even these direct-perception theorists will allow, of course, that perception of the mental states of other people is grounded in awareness of their behaviour. Yet this is widely agreed to be unnecessary in one's own case. One doesn't need to observe one's own movements, nor listen to one's own speech acts, in order to know what one is thinking. On the contrary, it is said that one can know this immediately and introspectively.

Carruthers (2011) provides an extended argument that this common-sense picture of self-knowledge is mistaken. On the contrary, knowledge of one's own thoughts is just as interpretive as is knowledge of the mental states of others. It draws on the same, or very similar, folk-psychological resources, only with one's 'mind-reading faculty' directed toward oneself rather than toward other people. And the same sorts of informational channels are relied upon in each case. Of course, the *data* utilized by the mind-reading system can differ in the first person. In particular, the system has access to the thinker's visual imagery, inner speech, and other sensory-like episodes, whereas it has no such access to the visual imagery or inner speech of other people (except indirectly, via their overt verbal reports). But note that this is access to sensory-based or sensory-like mental events, not to the underlying non-sensory thoughts. Moreover, the movement from awareness of one's own inner speech to the propositional attitudes thereby manifested is just as interpretive as is listening to the speech of another person.

The relationship between inner speech and thought requires some additional comment. Our best theory of inner speech is that it results from attention directed at a so-called 'forward model' of the predicted sensory consequences of the motor instructions for a specific speech-act (Carruthers, 2011; Tian and Poeppel, 2012; Scott, 2013). Whenever actions in general are initiated (including speech actions), an 'efferent copy' of the motor instructions is created and used to generate a predictive model of the likely sensory consequences of the movement. (In cases of overt action, these are compared with afferent sensory feedback and used to make fine-grained online adjustments to one's movements as required; Jeannerod, 2006.) In the case of *inner* speech, motor instructions are created as normal, issuing in a forward model, but the outgoing signals to the muscles themselves are suppressed. Since motor instructions are low-level non-conceptual representations, any semantic information deriving from the thought-to-be-expressed will have been left behind in the sensory forward-model. The latter therefore needs to be received as input by the language comprehension system (included in which is the mind-reading system, which handles pragmatics) and processed and interpreted in something like the normal way.

If inner speech, like the speech of other people, needs to be interpreted, however, then how is it that we never hear our own inner speech as ambiguous, nor puzzle about what it might mean? For these are frequent occurrences when listening to the speech of others. The

answer has to do with the role of *accessibility* of conceptual and syntactic structures in normal speech interpretation (Sperber and Wilson, 1995).[3] Speech interpretation is strongly biased by context, especially by prior conversational context. Concepts and structures that are still easily accessible (remaining in a partially activated state) are prioritized. For example, one will normally pick as the intended referent for a pronoun the individual who was most recently mentioned in the discourse (and whose singular concept is thus most readily accessible). But when the speech in question is one's own inner speech, the relevant concepts and syntactic structures will have been in a fully activated state just fractions of a second prior to the onset of the interpretive process. The latter will thus be strongly biased, albeit biased veridically, toward the intended interpretation.

If self-knowledge results from self-directed mind-reading, then a number of predictions can be made. One is that there should be no dissociations (in either direction) between capacities for self-knowledge and capacities for other-knowledge. That is, there should be no people in whom self-knowledge remains intact while other-knowledge is damaged. Nor should there be any people in whom other-knowledge remains intact while self-knowledge is damaged. Moreover, the same cortical networks should be implicated in each. Carruthers (2011) examines alleged cases of dissociation in autism and schizophrenia, as well as data from brain imaging experiments. He argues that none of the claimed dissociations turns out to be real. On the contrary, deficits in other-knowledge seem always to be paired with similar deficits of self-knowledge, and the brain networks implicated in both forms of knowledge are the same.

If self-knowledge of thoughts isn't direct, but results rather from self-directed mind-reading, then a further prediction can be made. This is that there should be distinctive patterns of error in people's claims about their own thoughts, mirroring the ways in which we can be misled about the thoughts of others. Care needs to be taken to delineate this prediction precisely, however. For an introspection-theorist might grant that there is nothing special about one's knowledge of one's own *past* thoughts (Nichols and Stich, 2003). It may

[3] Note that the relation of accessibility in play here is much broader than that involved in access *consciousness*, and applies within and between cognitive systems quite generally. For example, one syntactic structure may be more accessible within the language faculty because it was more recently activated, and is thus more easily *re*-activated.

well be that no long-term memories of one's own thought processes are generally kept, so that knowledge of one's past thoughts must depend on interpretation of what one does remember, namely one's past circumstances and behaviour. The crucial data therefore concern errors about one's own current or very recent thoughts.

Carruthers (2011) reviews a number of bodies of evidence suggesting that people do *not* have introspective access to their own thoughts, specifically their own current beliefs. One set derives from the 'self-perception' framework in social psychology, which has been extensively investigated (Bem, 1972; Albarracín and Wyer, 2000; Barden and Petty, 2008). For example, people duped into nodding while listening to a message (ostensibly to test the headphones they are wearing) report greater agreement with the content of the message, whereas those induced to shake their heads while listening report reduced agreement (Wells and Petty, 1980). This suggests that people interpret their own behaviour and modify their reports accordingly. Moreover, these effects can be made to reverse if the messages are unpersuasive — in this case nodding *decreases* belief in the message rather than increasing it, suggesting that nodding is interpreted as agreement with one's own internally accessible reactions, like thinking to oneself in inner speech, 'What an idiot!' (Briñol and Petty, 2003).

Similarly, right-handed people who write statements about themselves with their right hands thereafter express greater confidence in the truth of those statements when re-reading them than do those who write using their left hands (*ibid.*). It seems the shaky writing in the latter case is interpreted as a sign of hesitancy. And indeed, third parties who are asked to judge the degree of confidence of the writer from the handwriting samples alone display the same effect, and to the same extent.

Carruthers (2011) also discusses evidence from the counter-attitudinal essay paradigm in psychology, which has likewise been heavily investigated (Festinger, 1957; Elliot and Devine, 1994; Simon, Greenberg and Brehm, 1995; Gosling, Denizeau and Oberlé, 2006). People who are manipulated into feeling that they have made a free choice to write an essay arguing for the opposite of what they believe will thereafter shift their reports of their beliefs quite markedly — moving, for example, from being strongly opposed to a rise in college tuition to being neutral or mildly positive. This is known not to be an effect of argument quality, and people shift their reports without being aware of having done so, and without there being any prior change in

the underlying belief. Rather, what people are doing is managing their own emotions: they are making themselves feel better about what they have done, having had the sense that they had done something bad (indeed, people who are duped into thinking that they have caused harm through their freely undertaken advocacy of what they actually believe will also shift their reports of their beliefs to make themselves feel better; Scher and Cooper, 1989). But one would think that a direct question about what one believes would activate that belief and make it available for introspection, if such a thing were possible at all. Yet plainly people aren't aware of their beliefs at the time when they answer the query. Otherwise they would be aware that they are lying and would feel worse, not better (which is what they actually do).

Of course it is possible for a defender of introspection to respond to this (and voluminous other) evidence by allowing that people *sometimes* rely on indirect methods when ascribing thoughts to themselves (Rey, 2013). This is consistent with claiming that people are also capable of directly accessing their thoughts, perhaps in other circumstances or in other cases. Aside from being *ad hoc*, however, this manoeuvre makes no concrete predictions — it tells us nothing about the circumstances in which people will go wrong. And by the same token, it is incapable of explaining the patterning in the data. Why should errors of self-attribution emerge especially in cases where behavioural evidence might also mislead an outside observer, as well as in cases where people are motivated (unconsciously) to say something other than they believe? If people were genuinely capable of introspecting their thoughts, then it is remarkable that such abilities should happen to break down here and not elsewhere.

Following extensive discussion, Carruthers (2011) concludes from these and other arguments that our access to our own thoughts is always interpretive, no different in principle from our access to the thoughts of other people. While self-knowledge can rely on sensory data not available in the case of other people (including one's own visual imagery and inner speech), and while various factors may make self-knowledge more reliable than other-knowledge, both are equally indirect and interpretive in nature. In consequence, if conscious thoughts are those that one has immediate introspective knowledge of, then it follows that there are no such things.

3. Sensory-Based Broadcasting

If one's thoughts aren't higher-order access-conscious (that is, immediately knowable through introspection), then perhaps they are first-order access-conscious. Perhaps thoughts can be 'globally broadcast' and made available to a wide range of systems in the mind-brain. (The list of systems involved would normally be said to include those for drawing inferences, for forming memories, for generating affective reactions, for planning and decision making, and for verbal report.) One immediate problem with such a proposal, however, is that it seemingly conflicts with the confabulation data discussed in Section 2. For if one's thoughts are globally broadcast and made available to the systems responsible for verbal report, then one might think it should be a trivial matter to produce direct reports of them.

Perhaps this objection isn't devastating. It may be that once one's beliefs have been activated by a query, for example, they are globally broadcast and made *available* for verbal report; but the processes that plan and determine the nature of those reports can be unconscious ones. Perhaps other information besides the globally broadcast belief can be drawn on when formulating a report; and perhaps normal instances of speech production can be influenced (unconsciously) by a variety of motivational and other factors. In that case the belief might count as conscious at the same time that one misreports it, and while one is unaware that one is misreporting it. This combination of views might strike one as quite puzzling. But perhaps it isn't incoherent. In any case it will be fruitful to evaluate the claim that thoughts can be first-order access-conscious on its own merits.

Contradicting such a claim, Carruthers (2015a) argues that all access-conscious mental states are sensory-based, in that their conscious status constitutively depends upon some or other set of content-related sensory components (that is, perceptual states or mental images in one sense-modality or another). Amodal concepts can be bound into the content of these access-conscious states, however. Thus one doesn't just imagine colours and shapes, but a palm tree on a golden beach, for example. Here the concepts PALM TREE, GOLDEN, and BEACH are bound into the visual image in the same way (and resulting from the same sorts of interactive back-and-forth processing) as they are when one sees a scene as containing a palm tree on a golden beach. But the access-conscious status of these concepts is dependent on the presence of the sensory representations into which they are bound.

It is worth saying more about how conceptual representations can be bound into sensory or sensory-like states, since this will help us to see how one can perceive the thoughts of other people (as argued briefly in Section 2) and of ourselves (as will become important in Section 4). We know that visual processing, for example, takes place in a distributed fashion, with colour being processed separately from shape, and both being processed independently of movement. Yet each of these separate properties can be bound together into a single percept of, say, a round red object (a tomato) rolling along a surface (or in other cases, an integrated visual image of such an event). A central organizing principle in the binding process are so-called 'object files' (Pylyshyn, 2003). These are like indexical links to an object ('*That thing…*') to which property information (colour, shape, and the rest) can be attached.

Carruthers (2015a) then argues that the best account of seeing *as* (where the round red object is seen *as* a tomato, for instance) is that category information can be bound into these object files and globally broadcast along with them, constituting a single conscious visual percept. For the competing view would have to be that there are two distinct conscious events: one is a perceptual object file ('*That*: round red rolling thing') whereas the other is a perceptual judgment ('*That*: TOMATO'). Notice, however, that such an account faces a new version of the binding problem. For it fails to explain what secures the coincidence of reference of the two indexicals, making it the case that one sees the round red rolling thing as the tomato, rather than something else in the visual field.[4]

When we turn to speech perception (and by extension, inner speech), the relevant organizing principle is the *event file*. (An object-file structure is unlikely to work here, since the only relevant object would be the speaker. But one can understand speech, and bind it into a single interpreted utterance, without knowing or otherwise perceiving the identity of the speaker.) Speech is segmented into distinct events (generally sentences), with multiple properties drawn from

[4] It is important to notice that the view proposed here, that conceptual information can be bound into perceptual and imagistic states and globally broadcast along with them, is perfectly consistent with claiming that perceptual *systems* are distinct from conceptual ones. One can claim that there are cortical networks specialized for processing information from the retina, for example, while also allowing that those networks interact with amodal conceptual ones, and that globally broadcast visual representations can comprise both sorts of representation.

many different levels of processing bound into each event file. Thus one hears the tone of voice, the volume, and the accent with which people say things, while also hearing what they say, and often also the intent with which they say it (as when one hears someone as speaking ironically, for example). As a result, an auditory event file can have mental-state information bound into it.

Returning, now, to the main theme of this section: one argument for the view that all access consciousness depends upon sensory representations is an inference to the best explanation that brings together recent work on consciousness with recent work on working memory. The argument builds on the findings of Baars (1988; 2002; 2003), Dehaene (Dehaene *et al.*, 2006; Dehaene and Changeux, 2011; Dehaene, 2014), and others who have amassed a large and convincing body of data in support of the 'global broadcasting' or 'global workspace' theory of conscious experience. Across a wide variety of unconscious forms of perception there can be local reverberating activity in both mid-level and high-level sensory cortices. (In the case of vision, these include the occipital cortex and posterior temporal cortex.) Stimuli in such cases can be processed all the way up to the conceptual level while remaining unconscious, which can give rise to semantic priming effects. But when this activity is targeted by attention the percepts become conscious, and there is widespread coordinated activity linking it also to frontal and parietal cortices.[5]

Everyone agrees that attention can be a major determinant of consciousness. Carruthers (2015a) goes further and argues that it is necessary and (with other factors) sufficient for consciousness. While some have claimed that gist perception and/or background-scene perception is conscious in the absence of attention, recent studies have shown that this is incorrect: such properties merely require comparatively little attention to be consciously perceived (Cohen, Alvarez and Nakayama, 2011; Mack and Clarke, 2012). Moreover, the neural mechanisms underlying attention are increasingly well understood (Baluch and Itti, 2011; Bisley, 2011). A top-down attentional network links the dorsolateral prefrontal cortex, the frontal eye-fields, and the intraparietal sulcus. The 'business end' of the system is the latter,

5 It should be noted that both Baars and Dehaene take for granted that thoughts as well as experiences can be globally broadcast (Baars, Franklin and Ramsøy, 2013; Dehaene, 2014). Yet they offer no positive evidence for such a view. A reasonable inference is that they, too, fall prey to the illusion of conscious thought, and are merely endorsing a common-sense extension of their scientific findings that strikes them as obvious.

which projects both boosting and suppressing signals to targeted areas of mid-level sensory cortices (see also Prinz, 2012). At the same time there is a bottom-up attentional network (sometimes called the 'saliency network') linking regions of the right ventrolateral parietal cortex and right ventrolateral prefrontal cortex, which then interact with the top-down system through the anterior cingulate cortex (Corbetta, Patel and Shulman, 2008; Sestieri, Shulman and Corbetta, 2010).

An extensive recent body of research on working memory suggests that this same attentional network, which is responsible for conscious perception, is also involved in our capacity to sustain and generate conscious representations endogenously, for purposes of conscious thinking and reasoning. For example, whenever brain imaging studies of working memory have been conducted using appropriate subtraction tasks, content-related activity in one or more sensory areas has been found (Postle, 2006; 2016; D'Esposito, 2007; Jonides *et al.*, 2008; Serences *et al.*, 2009; Sreenivasan, Sambhara and Jha, 2011). Moreover, this activity plays a causal role in the tasks in question, since transcranial magnetic stimulation (TMS) applied to these areas during the retention interval in working memory tasks disrupts performance (Herwig *et al.*, 2003; Koch *et al.*, 2005).[6] Notice, in addition, that most working memory tasks *could* be solved purely amodally, if such a thing were really possible — keeping numbers, words, or concepts active in the global workspace. Yet this doesn't seem to happen.

An inference to the best explanation enables us to combine and unify these two bodies of research, thereby detailing the mechanisms that underlie the stream of consciousness quite generally. Attentional signals directed at mid-level sensory areas are necessary for contents to enter working memory (thereby becoming conscious), as well as for conscious perception. And then if working memory is the system that underlies conscious forms of reasoning and decision making, as many in the field believe (Evans and Stanovich, 2013; Carruthers, 2015a), we can conclude that all conscious thinking is sensory-based.

It remains possible, of course, that there is, in addition to a sensory-based working memory system, an amodal (non-sensory) workspace

[6] Transcranial magnetic stimulation (TMS) involves targeting specific regions of the cortex with a series of weak magnetic pulses, thereby introducing 'noise' into the processing being conducted in those regions.

in which thoughts and propositional attitudes can figure consciously. However, we have no evidence of any form of global broadcasting that isn't tied to sensory cortex activity. Nor do we have evidence of an attentional network with the right 'boosting and suppressing' properties targeted at the anterior and medial temporal cortex or prefrontal cortex, which is what would be needed if amodal thoughts were to be globally broadcast. Of course, absence of evidence isn't evidence of absence by itself. But Carruthers (2015a) discusses a number of lines of argument that count strongly against the competing proposal outlined here. What follows is a sketch of one of them.

Suppose there is some sort of workspace in which amodal (non-sensory) thoughts — judgments, goals, decisions, intentions, and the rest — can become active and be conscious. What would one predict? One would surely expect that variance in the properties of this workspace among people would account for a large proportion of people's variance in fluid general intelligence, or fluid g. For it is conscious forms of thinking and reasoning that are believed to underlie our capacity to solve novel problems in creative and flexible ways, which are precisely the abilities measured by tests of fluid g.

In fact, there are now a great many studies examining the relationship between working memory and fluid g (Conway, Kane and Eagle, 2003; Colom *et al.*, 2004; 2008; Cowan *et al.*, 2005; Kane, Hambrick and Conway, 2005; Unsworth and Spillers, 2010; Redick *et al.*, 2012; Shipstead *et al.*, 2014). Generally, variance in the former overlaps with the latter somewhere between 0.6 and 0.9 (that is to say, the relationship between the two seems to lie somewhere between *very strong* and *almost identical*). Many have thus come to regard working memory as the cognitive system or mechanism that is responsible for fluid g (which is itself a purely statistical construct, of course, being the underlying common factor calculated from a range of different types of reasoning task). And to the extent that other factors have been found to correlate with fluid g independently of working memory, the only one that has received robust support is speed of processing, which seems to be a low-level phenomenon (perhaps related to the extent of neural myelination).

It may be, of course, that standard tests of working memory tap into both the sensory-based system and the supposed amodal thought-involving system. But in that case one would predict that, as tests of working memory become more and more sensory in character (requiring one to keep in mind or manipulate un-namable shapes or shades of colour, for example), the extent of the overlap with fluid g

should decrease. For these tests of purely sensory working memory would fail to include any measure of the variance in amodal thinking abilities that would (by hypothesis) account for a large proportion of our flexible general intelligence. But this seems not to be the case. Low-level sensory tasks overlap with fluid g just as strongly (if not more strongly) than do concept-involving ones (Unsworth and Spillers, 2010; Burgess *et al.*, 2011; Redick *et al.*, 2012; Shipstead *et al.*, 2012; 2014). Moreover (and just as the sensory-based account would predict) measures of sensory attentional control (using such tests as the anti-saccade task or the flankers task) themselves predict capacities for general intelligence quite strongly (Unsworth and Spillers, 2010; Shipstead *et al.*, 2012; 2014).[7]

In addition, there is a separate body of evidence that pushes toward the same conclusion. This derives from studies that have presented people with a range of different sensory discrimination tasks (Acton and Schroeder, 2001; Deary *et al.*, 2004; Meyer *et al.*, 2010; Voelke *et al.*, 2014). Participants might be asked to order a series of colour-chips by shade, arrange a series of lines by length, arrange a set of tones by pitch, order a set of identical-looking objects by manually feeling their weight, and so on. From these measures one can compute an underlying common factor (just as one does when computing fluid g from a range of reasoning tasks). While it is unclear exactly what this common factor represents, it seems likely that it has to do with capacities for sensory attention and purely-sensory working memory. Across studies, it has been found that this underlying factor overlaps with fluid g between 0.6 and 0.9 (note that this is the same as the extent of overlap between working memory and fluid g). Since there will be executive and memory-search components of working memory that make no contribution to these sensory discrimination tasks, we can conclude pretty confidently that there is no variance in general intelligence remaining to be explained by the hypothesized workspace for conscious amodal thinking and reasoning.

Carruthers (2015a) argues on these and other grounds that only mental states that have a sensory-based format (such as visual or auditory imagery) are capable of becoming first-order access-

7 The anti-saccade task requires participants to saccade *away from* a suddenly appearing visual cue, rather than towards it, which is what one naturally does. The flankers task requires one to indicate the direction of a central arrow that is flanked by others that can be either congruent or incongruent in their direction (with the latter being more difficult).

conscious. When taken together with the conclusion of Section 2, it follows that amodal thoughts are neither first-order access-conscious nor higher-order access-conscious. All thoughts must therefore do their work unconsciously — among other things, helping to direct attention and manipulate sensory-based representations in working memory.

4. Whence the Illusion?

The evidence suggests, then, that there are no such things as conscious thoughts. On the contrary, all conscious thinking and reasoning requires a sensory-based format, involving imagery of one sort or another. Amodal thoughts exist, of course. We make judgments, access memories and beliefs, form and act on goals and intentions, and so on. But such thoughts are always unconscious. They mostly do their work downstream of the conscious contents of working memory. They may be evoked into activity by conscious states, perhaps, but they enter into processes of reasoning and decision making that fall outside the content of working memory, and are *un*conscious.

There are in addition, of course, processes of reasoning that take place *in* working memory, and are conscious. These are so-called 'System 2' inferential processes (Kahneman, 2011; Evans and Stanovich, 2013). But they operate over sentences of inner speech, visual imagery, and other sensory-based contents. System 2 processes do not, therefore, include amodal thoughts (or at least, not on the account being defended here). Furthermore, unconscious thoughts also work behind the scenes generating and controlling the sensory-based contents that figure in working memory and the stream of consciousness itself (Carruthers, 2015a).

What remains, however, is a puzzle: if there are no conscious thoughts, then why does almost everyone believe that there are? How do we come to be under the illusion of conscious thought? This is the question to be addressed here.

A number of different factors need to be combined together to construct an adequate explanation. One is a point discussed briefly in Section 2. This is that the central role played by *accessibility* of concepts and syntactic structures in the interpretation of speech (whether internal or external) means that one fails to notice ambiguities in one's own inner speech, and ensures that the latter hardly ever strikes one as puzzling or incomprehensible. This is because the relevant conceptual and syntactic structures of the thought-to-be-expressed in the

rehearsed speech-act will have been active immediately prior to the start of the comprehension process, strongly biasing the latter. A single interpretation almost always wins out as a result, and it does so smoothly and swiftly (just as it does in connection with one's own overt speech).

A second factor has also already been mentioned. This is that we often seem to *see* or *hear* what people are thinking (Carruthers, 2015b). This is most obvious in connection with speech. If someone stops one in the street and asks the way to the Adventist church, one may *hear* her as *wanting to know* where the church is. From one's own subjective perspective, it is not that one *first* hears the sounds that she makes and then figures out what she wants (something like this may well be going on unconsciously, of course). Rather, understanding is seemingly immediate, and a mental-state attribution comes bound into the content of the sound stream. Similarly, if one asks a work colleague when a scheduled meeting begins and he replies, 'It starts in ten minutes', one *hears* him as *judging*, or as *believing*, that the meeting begins then. Again a thought attribution is bound into the content of what one hears. Likewise for visual perceptions of someone's behaviour: in many cases one's experience is imbued with mental-state content. Thus one might *see* someone as *trying to open* a door, for example (as she struggles with the key in the lock), or one might *see* someone as *deciding to stop* to pick up a piece of litter (as he pauses and begins to bend down towards it).

Something similar is true of one's own inner speech. One can hear oneself as *wondering* whether it is time to leave for the bus, or as *judging* that it is. Representations of one's own thoughts are thus bound into the contents of one's reflective thinking, in such a way that one *experiences* oneself *as* entertaining those thoughts, seemingly immediately, and without engaging in any form of inference or self-interpretation. Likewise in connection with visual forms of thinking, using visual imagery. When one manipulates images of items of luggage while looking into the trunk of one's car one might experience oneself was *wondering* how those items will fit, or as *deciding* to push the large suitcase to the back. Again one's experience comes imbued with thought-attributions bound into it. Indeed, one's thoughts can strike one as being right there among the contents of one's auditory or visual imagistic experience.

The *experience* of deciding something is not the same thing as *deciding*, of course. The former is meta-representational, whereas the latter is not. So there will be two events here, having quite different

contents and causal roles. Moreover, on the view outlined in Section 2, the experience of deciding may-or-may-not correctly represent the presence of a corresponding decision. One can experience oneself as deciding something when really one is not, or while one is actually deciding something different.

It should be noted, however, that not all inner speech (nor other forms of imagery) is experienced in terms of some specific attitude. This will depend on whether the right sorts of contextual and other cues are present to enable the mind-reading system to determine an attribution, and on the speed with which it is able to do so. And in fact, one often experiences oneself as entertaining what Cassam (2014) calls 'a passing thought' — that is, a propositional content that isn't the object of any particular mental attitude. For example, one might report an episode in which hears oneself saying in inner speech, 'Time to go home', by saying, 'I was thinking *about* whether it is time to go home'. (Note that this isn't the same as saying, 'I was *asking myself* whether it is time to go home'. Nor is it the same as saying, 'I was *wondering* whether it is time to go home'. These attribute a particular mental attitude, that of asking a question, or of wanting to know something.) One was aware of a thought with the *content* that it is time to go home, that is all.

It is easy to explain why one should have the impression that one often knows one's own thoughts immediately and introspectively, then. For that is how one seemingly experiences them. Moreover, it is easy to understand why one should have the impression that one's thoughts in these circumstances are first-order access-conscious. For one fails to have any impression of *distance* between the thoughts themselves and the contents of one's conscious experience. And yet of course the thoughts that one attributes to oneself in these circumstances will seemingly be available to be remembered, to inform one's decision making, and to issue in verbal reports. However, why should one have the impression that one's access to one's own thoughts differs in *kind* from one's access to the thoughts of other people? For one's access to other people's thoughts is often just as phenomenally immediate. How, then, are we to explain the strength of the intuition of a self–other asymmetry?

One horn of the asymmetry is straightforward. For of course it is part of common sense that our access to the thoughts of other people is interpretive and mediated via perception of their circumstances and behaviour, despite the seeming phenomenal immediacy of many instances of thought-attribution. But what of the other horn? Why do

we never challenge the seeming immediacy of our access to our own thoughts? The answer, I suggest, is built into the structure of the mind-reading system itself. Specifically, the latter employs a tacit rule of interpretation, which is used in the third-person as well as in the first. This is that if someone *thinks* they are undergoing a certain mental state, then so they are. In fact, I suggest that something resembling Cartesian certainty about self-knowledge is built into our folk psychology. Not many people actually (explicitly) believe this any longer, of course (at least not once they have had some exposure to cognitive science). But that is not the idea. The claim is rather that Cartesian certainty about current mental events is implicit in a mind-reading inference-rule, which mandates that one move immediately from the belief that one *thinks* one is in mental state M to the conclusion that one therefore *is* in M. I shall refer to this as 'the Cartesian inference-rule'.[8]

One argument for such a view is an inference to the best explanation of the seeming universality of Cartesian beliefs across cultures and historical eras. As Carruthers (2011) reports (drawing partly on personal communications from experts in the relevant fields), whenever people in pre-scientific cultures have reflected on the nature of self-knowledge, they have assumed that their access to their own current thoughts is direct and immediate. Not only is this true in the history of Western philosophy, but it is also true of ancient China, the Buddhist tradition, and even the ancient Aztecs. If one rejects such views (as I have done) and argues that one's access to one's own thoughts is always indirect and interpretive, then this presents a puzzle. Why has almost everyone across cultures and times believed the opposite? The puzzle is removed if some version of the Cartesian assumption is built into the fabric of an innately channelled mind-reading system (for the existence of which there is now a significant body of evidence; see Barrett *et al.*, 2013; Carruthers, 2013b).

Such a claim is surely ripe for experimental testing. But any such tests should be designed to use indirect measures, rather than asking people to make explicit judgments about imagined scenarios (as did Kozuch and Nichols, 2011). Or if direct measures are used, the tests

8 Carruthers (2011) argues in addition that the converse rule — 'if someone is in mental state M, then they believe they are in mental state M' — is also tacitly encoded in the processing principles of the mind-reading system. This is why there was, initially, such vigorous resistance to the idea of unconscious mentality.

should be speeded or conducted under cognitive load. For the hypothesis isn't that people explicitly believe in Cartesian access to their own thoughts (on the contrary, educated people today probably don't). It is rather that an implicit processing-rule tantamount to such a belief governs the online processing of the mind-reading system. Anecdotally, however, it does seem that stimuli designed to violate the direct-access assumption generally strike one as somehow *weird*. Even after extensive reflection, and having written books on the subject, sentences such as 'I believe I have just decided to leave for the bus, but I haven't really decided that', or 'I have just decided to leave for the bus, but what is my evidence that I have just decided that?', strike me initially as being strange to the point of being almost ill-formed.

Why would the mind-reading system employ such a tacit principle of interpretation? In short, because it greatly simplifies the process of other-interpretation, probably without any loss of reliability. Let us take these points in turn. Much of the work of the mind-reading system lies in assisting one to interpret the speech of other people. It helps one to figure out which object someone is referring to in a context where indexicals or pronouns are employed. It helps one to determine whether the speech act is literal, ironic, joking, or whatever. And in the case of assertoric discourse, it helps one to judge whether the person is being honest or is attempting to deceive, and in evaluating their degree of certainty. Moreover, much of people's ordinary discourse concerns their own (and other people's) mental states. People talk about what they want, what they feel, what they think, and so on. These are complex matters. Yet for the most part comprehension happens smoothly and in real time. If the mind-reading system did *not* employ the Cartesian inference-rule, then in addition to figuring out whether the speaker is asserting something literally and honestly when she says she is in mental state M, the system would also need to determine whether the speaker is interpreting her own behaviour and internal cues correctly. This would add a whole extra layer of complexity, slowing down the interpretive process considerably. And there would probably be no gain in reliability to compensate, as I will now try to show.

Much of the data required to evaluate whether someone is interpreting herself correctly is simply not available. One almost never knows what someone is, or has been, visually imagining, nor the sentences that have been rehearsed in her inner speech. Other evidence would be costly to retrieve from long-term memory, such as

relevant behaviour from the person's past. Moreover, whatever evidence one *can* retrieve is likely to be fragmentary and incomplete, which provides an additional source of error. It is now a familiar point in cognitive science that simple heuristics can outperform more elaborate and information-hungry principles of judgment, not just in speed but also in reliability (Gigerenzer *et al.*, 1999). For if the data required for the operation of the information-hungry principle are incomplete and unrepresentative, then this may introduce errors that don't get made by the simpler heuristic system.

Sometimes, of course, we have behavioural evidence that conflicts with what someone says about her mental state. Think, for example, of a person who is red in the face and banging the table aggressively while yelling 'I am *not* angry!' In this case it *might* be useful to think that the person has misinterpreted her own state. So this is a case where the Cartesian inference-rule will close off possibilities that it might actually be fruitful to consider. But even here it is doubtful whether anything important is lost for most practical purposes. For one can (and does) easily attribute the discrepancy in the person's behaviour to disingenuousness. One can think that the person is trying to mislead her audience, and is not reporting her emotional state honestly. This enables one to form expectations based on an attribution of anger while dismissing the person's verbal statement, but it does so while retaining the simplifying Cartesian inference-rule.

If sceptical doubts are raised, then, about the directness of one's attributions of thoughts to oneself, they are apt to be immediately silenced, or closed off, through an application of the Cartesian inference-rule. If one is apt to treat 'I believe I am in mental state M' as entailing 'I am in mental state M', then the question whether one might *take* oneself or *interpret* oneself to be in M without really being so will never even arise. And if such a possibility *is* raised, it will strike one that it should immediately be rejected. By the same token, the suggestion that one might know of one's own current thoughts and attitudes in the same way that one knows of the attitudes of other people — by interpreting sensory cues of one sort or another — will strike one as inherently absurd.

In short, then, the reason why we are under the illusion of conscious thought is that our access to our own thoughts is seemingly direct and perception-like (as is our access to the thoughts of other people, on many occasions). But (in stark contrast with our awareness of others' thoughts) we are prevented from recognizing the interpretive, non-immediate, character of our access to our own thoughts by the

inferential structure of the mind-reading system that provides us with that access.

5. Conclusion

I have argued that amodal (non-sensory) thoughts such as beliefs, goals, and decisions are never conscious in either the first-order or the higher-order access sense. Such thoughts are never globally broadcast and made available to a wide range of systems in the mind-brain. Nor are they capable of being known directly and without interpreting sensory cues. On the contrary, amodal thoughts operate beneath the level of awareness, influencing both overt and covert forms of action, and one's knowledge of them results from interpreting sensory-based cues of various sorts (primarily overt behaviour and mental imagery). Yet the interpretive process is swift and generally reliable, to the point where one routinely *experiences* oneself *as* entertaining thoughts of various kinds. Moreover, a Cartesian-like inference-rule built into the structure of the interpreting system (the mind-reading faculty) blocks sceptical doubts about one's knowledge of one's own states of mind, while making it seem as if one's access to one's own thoughts differs in *kind* from one's access to the thoughts of other people.

References

Acton, G. & Schroeder, D. (2001) Sensory discrimination as related to general intelligence, *Intelligence*, **29**, pp. 263–271.

Albarracín D. & Wyer, R. (2000) The cognitive impact of past behavior: Influences on beliefs, attitudes, and future behavioral decisions, *Journal of Personality and Social Psychology*, **79**, pp. 5–22.

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

Baars, B. (2002) The conscious access hypothesis: Origins and recent evidence, *Trends in Cognitive Sciences*, **6**, pp. 47–52.

Baars, B. (2003) How brain reveals mind: Neuroimaging supports the central role of conscious experience, *Journal of Consciousness Studies*, **10**, pp. 100–114.

Baars, B., Franklin, S. & Ramsøy, T. (2013) Global workspace dynamics: Cortical 'binding and propagation' enables conscious contents, *Frontiers in Psychology*, **4**, art. 200.

Baluch, F. & Itti, L. (2011) Mechanisms of top-down attention, *Trends in Neurosciences*, **34**, pp. 210–224.

Barden, J. & Petty, R. (2008) The mere perception of elaboration creates attitude certainty: Exploring the thoughtfulness heuristic, *Journal of Personality and Social Psychology*, **95**, pp. 489–509.

Barrett, H.C., Broesch, T., Scott, R., He, Z., Baillargeon, R., Wu, D., Bolz, M., Henrich, J., Setoh, P., Wang, J. & Laurence, S. (2013) Early false-belief under-

standing in traditional non-Western societies, *Proceedings of the Royal Society B: Biological Sciences*, **280**, 20122654.
Bem, D. (1972) Self-perception theory, *Advances in Experimental Social Psychology*, **6**, pp. 1–62.
Bisley, J. (2011) The neural basis of visual attention, *Journal of Physiology*, **589**, pp. 49–57.
Briñol, P. & Petty, R. (2003) Overt head movements and persuasion: A self-validation analysis, *Journal of Personality and Social Psychology*, **84**, pp. 1123–1139.
Burgess, G., Gray, J., Conway, A. & Braver, T. (2011) Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span, *Journal of Experimental Psychology: General*, **140**, pp. 674–692.
Carruthers, P. (2000) *Phenomenal Consciousness*, New York: Cambridge University Press.
Carruthers, P. (2011) *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford: Oxford University Press.
Carruthers, P. (2013a) On knowing your own beliefs: A representationalist account, in Nottelmann, N. (ed.) *New Essays on Belief: Structure, Constitution and Content*, New York: Palgrave Macmillan.
Carruthers, P. (2013b) Mindreading in infancy, *Mind & Language*, **28**, pp. 141–172.
Carruthers, P. (2015a) *The Centered Mind: What the Science of Working Memory Shows Us about the Nature of Human Thought*, Oxford: Oxford University Press.
Carruthers, P. (2015b) Perceiving mental states, *Consciousness and Cognition*, **36**, pp. 498–507.
Carruthers, P. & Veillet, B. (2011) The case against cognitive phenomenology, in Bayne, T. & Montague. M. (eds.) *Cognitive Phenomenology*, Oxford: Oxford University Press.
Cassam, Q. (2014) *Self-Knowledge for Humans*, Oxford: Oxford University Press.
Cohen, M., Alvarez, G. & Nakayama, K. (2011) Natural-scene perception requires attention, *Psychological Science*, **22**, pp. 1165–1172.
Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M. & Kyllonen, P. (2004) Working memory is (almost) perfectly predicted by g, *Intelligence*, **32**, pp. 277–296.
Colom, R., Abad, F., Quiroga, A., Shih, P. & Flores-Mendoza, C. (2008) Working memory and intelligence are highly related constructs, but why?, *Intelligence*, **36**, pp. 584–606.
Conway, A., Kane, M. & Engle, R. (2003) Working memory capacity and its relation to general intelligence, *Trends in Cognitive Sciences*, **7**, pp. 547–552.
Corbetta, M., Patel, G. & Shulman, G. (2008) The reorienting system of the human brain: From environment to theory of mind, *Neuron*, **58**, pp. 306–324.
Cowan, N., Elliott, E., Saults, J.S., Morey, C., Mattox, S., Hismjatullina, A. & Conway, A. (2005) On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes, *Cognitive Psychology*, **51**, pp. 42–100.
Deary, I., Bell, P., Bell, A., Campbell, M. & Fazal, N. (2004) Sensory discrimination and intelligence: Testing Spearman's other hypothesis, *American Journal of Psychology*, **117**, pp. 1–18.

Dehaene, S. (2014) *Consciousness and the Brain*, New York: Viking Press.
Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J. & Sergent, C. (2006) Conscious, preconscious, and subliminal processing: A testable taxonomy, *Trends in Cognitive Sciences*, **10**, pp. 204–211.
Dehaene, S. & Changeux, J.-P. (2011) Experimental and theoretical approaches to conscious processing, *Neuron*, **70**, pp. 200–227.
D'Esposito, M. (2007) From cognitive to neural models of working memory, *Philosophical Transactions of the Royal Society B*, **362**, pp. 761–772.
Elliot, A. & Devine, P. (1994) On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort, *Journal of Personality and Social Psychology*, **67**, pp. 382–394.
Evans, J. & Stanovich, K. (2013) Dual-process theories of higher cognition: Advancing the debate, *Perspectives on Psychological Science*, **8**, pp. 223–241.
Festinger, L. (1957) *A Theory of Cognitive Dissonance*, Palo Alto, CA: Stanford University Press.
Gallagher, S. (2001) The practice of mind: Theory, simulation, or primary interaction?, *Journal of Consciousness Studies*, **8** (5–7), pp. 83–107.
Gigerenzer, G., Todd, P. & the ABC Research Group (1999) *Simple Heuristics that Make Us Smart*, Oxford: Oxford University Press.
Gosling, P., Denizeau, M. & Oberlé, D. (2006) Denial of responsibility: A new mode of dissonance reduction, *Journal of Personality and Social Psychology*, **90**, pp. 722–733.
Graziano, M. (2013) *Consciousness and the Social Brain*, Oxford: Oxford University Press.
Herwig, U., Abler, B., Schönfeldt-Lecuona, C., Wunderlich, A., Grothe, J., Spitzer, M. & Walter, H. (2003) Verbal storage in a premotor-parietal network: Evidence from fMRI-guided magnetic stimulation, *NeuroImage*, **20**, pp. 1032–1041.
Hutto, D. (2004) The limits of spectatorial folk psychology, *Mind & Language*, **19**, pp. 548–573.
Jeannerod, M. (2006) *Motor Cognition*, Oxford: Oxford University Press.
Jonides, J., Lewis, R., Nee, D., Lustig, C., Berman, M. & Moore, K. (2008) The mind and brain of short-term memory, *Annual Review of Psychology*, **59**, pp. 193–224.
Kahneman, D. (2011) *Thinking, Fast and Slow*, New York: Farrar, Straus, and Grioux.
Kane, M., Hambrick, D. & Conway, A. (2005) Working memory capacity and fluid intelligence are strongly related constructs, *Psychological Bulletin*, **131**, pp. 66–71.
Koch, C., Oliveri, M., Torriero, S., Carlesimo, G., Turriziani, P. & Caltagirone, C. (2005) rTMS evidence of different delay and decision processes in a fronto-parietal neuronal network activated during spatial working memory, *NeuroImage*, **24**, pp. 34–39.
Kozuch, B. & Nichols, S. (2011) Awareness of unawareness: Folk psychology and introspective transparency, *Journal of Consciousness Studies*, **18** (11–12), pp. 135–160.
Mack, A. & Clarke, J. (2012) Gist perception requires attention, *Visual Cognition*, **20**, pp. 300–327.

Meyer, C., Hagmann-von Arx, P., Lemola, S. & Grob, A. (2010) Correspondence between the general ability to discriminate sensory stimuli and general intelligence, *Journal of Individual Differences*, **31**, pp. 46–56.
Nichols, A. & Stich, S. (2003) *Mindreading*, New York: Oxford University Press.
Noë, A. (2004) *Action in Perception*, Cambridge, MA: MIT Press.
Postle, B. (2006) Working memory as an emergent property of the mind and brain, *Neuroscience*, **139**, pp. 23–38.
Postle, B. (2016) How does the brain keep information 'in mind'?, *Current Directions in Psychological Science*, **25**, pp. 151–156.
Prinz, J. (2012) *The Conscious Brain*, New York: Oxford University Press.
Pylyshyn, Z. (2003) *Seeing and Visualizing*, Cambridge, MA: MIT Press.
Redick, T., Unsworth, N., Kelly, A. & Engle, R. (2012) Faster, smarter? Working memory capacity and perceptual speed in relation to fluid intelligence, *Journal of Cognitive Psychology*, **24**, pp. 844–854.
Rey, R. (2013) We are not all 'self-blind': A defense of a modest introspectionism, *Mind & Language*, **28**, pp. 259–285.
Rosenthal, D. (2005) *Consciousness and Mind*, Oxford: Oxford University Press.
Scher, S. & Cooper, J. (1989) Motivational basis of dissonance: The singular role of behavioral consequences, *Journal of Personality and Social Psychology*, **56**, pp. 899–906.
Schwitzgebel, E. (2002) A phenomenal, dispositional account of belief, *Noûs*, **36**, pp. 249–275.
Schwitzgebel, E. (2011) Knowing your own beliefs, *Canadian Journal of Philosophy*, **35**, pp. 41–62.
Scott, M. (2013) Corollary discharge provides the sensory content of inner speech, *Psychological Science*, **24**, pp. 1824–1830.
Serences, J., Ester, E., Vogel, E. & Awh, E. (2009) Stimulus-specific delay activity in human primary visual cortex, *Psychological Science*, **20**, pp. 207–214.
Sestieri, C., Shulman, G. & Corbetta, M. (2010) Attention to memory and the environment: Functional specialization and dynamic competition in human posterior parietal cortex, *The Journal of Neuroscience*, **30**, pp. 8445–8456.
Shipstead, Z., Redick, T., Hicks, K. & Engle, R. (2012) The scope and control of attention as separate aspects of working memory, *Memory*, **20**, pp. 608–628.
Shipstead, Z., Lindsey, D., Marshall, R. & Engle, R. (2014) The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control, *Journal of Memory and Language*, **72**, pp. 116–141.
Simon, L., Greenberg, J. & Brehm, J. (1995) Trivialization: The forgotten mode of dissonance reduction, *Journal of Personality and Social Psychology*, **68**, pp. 247–260.
Spaulding, S. (2016) On whether we can see intentions, *Pacific Philosophical Quarterly*, **98** (2), pp. 150–170.
Sperber, D. & Wilson, D. (1995) *Relevance: Communication and Cognition*, 2nd ed., Oxford: Blackwell.
Sreenivasan, K., Sambhara, D. & Jha, A. (2011) Working memory templates are maintained as feature-specific perceptual codes, *Journal of Neurophysiology*, **106**, pp. 115–121.
Tian, X. & Poeppel, D. (2012) Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation, *Frontiers in Human Neuroscience*, **6**, art. 314.

Tononi, G. (2008) Consciousness as integrated information: A provisional manifesto, *Biological Bulletin*, **215**, pp. 216–242.
Tononi, G., Boly, M., Massimini, M. & Koch, C. (2016) Integrated information theory: From consciousness to its physical substrate, *Nature Reviews Neuroscience*, **17**, pp. 450–461.
Unsworth, N. & Spillers, G. (2010) Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model, *Journal of Memory and Language*, **62**, pp. 392–406.
Voelke, A., Troche, S., Rammsayer, T., Wagner, F. & Roebers, C. (2014) Relations among fluid intelligence, sensory discrimination and working memory in middle to late childhood — A latent variable approach, *Cognitive Development*, **32**, pp. 58–73.
Wells, G. & Petty, R. (1980) The effects of overt head movements on persuasion, *Basic and Applied Social Psychology*, **1**, pp. 219–230.

Paper received October 2016; revised February 2017.