

Uwe Peters

Self-Knowledge and Consciousness of Attitudes

Abstract: *Suppose we know our own attitudes, e.g. judgments and decisions, only by unconsciously interpreting ourselves. Would this undermine the assumption that there are conscious attitudes? Carruthers (2011) has argued that if the mentioned view of self-knowledge is combined with either of the two most common approaches to consciousness, i.e. the higher-order state account (Rosenthal, 1997; 2005; Lycan, 1996; Carruthers, 2000) or the global workspace theory (Baars, 1988; Dehaene and Naccache, 2001), then the conjunction of these theories implies that there are no conscious attitudes. I shall show that Carruthers' argument against the existence of conscious attitudes doesn't succeed, and mention studies on autism and logical reasoning under cognitive load that suggest that there are conscious attitudes.*

Does the way we know our own attitudes, e.g. judgments and decisions, undermine the view that there are *conscious* attitudes?

Peter Carruthers (2011) has made a powerful case for the view that we know our own attitudes only by means of an unconscious process of self-interpretation.¹ He contends furthermore that if this view of self-knowledge is combined with either of the two most common approaches to consciousness, i.e. the higher-order state account (e.g. Rosenthal, 1997; 2005; Lycan, 1996; Carruthers, 2000) or the global workspace theory (e.g. Baars, 1988; Dehaene and Naccache, 2001),

Correspondence:
Email: uwe.peters@kcl.ac.uk

[1] Carruthers excludes 'sensorily-embedded judgments' (e.g. seeing *as*, or hearing *as*) and affective 'context-bound' desires and emotions from this claim (Carruthers, 2011, p. 10). The qualification should be kept in mind throughout this paper. See Carruthers (2011, pp. 2ff.) for details on these two exceptions.

then the conjunction of these theories implies that there are no conscious attitudes.²

In the following, I shall show that Carruthers' argument against the existence of conscious attitudes doesn't succeed, and mention studies on autism and logical reasoning under cognitive load that suggest that there are conscious attitudes.

Since Carruthers' case against conscious attitudes rests on his specific account of self-knowledge, I begin in Section 1 with a brief exposition of the latter before in Section 2 introducing and critiquing his case. In Sections 3 and 4, I then offer two arguments for conscious attitudes.

1. On Self-Knowledge of Attitudes

How do we know our own attitudes? It seems that while we know other people's mental states only by observation and interpretation of their behaviour, we know our own attitudes directly, i.e. without interpreting ourselves or circumstances (see, for example, Shoemaker, 1996; Nichols and Stich, 2003; Goldman, 2006). However, psychologists (e.g. Gopnik, 1993; Gazzaniga, 1995; Wegner, 2002; Wilson, 2002; Bem, 1967; Gopnik and Meltzoff, 1994; Frith and Happé, 1999) and philosophers (e.g. Ryle, 1948; Dennett, 1991; Lawlor, 2008; 2009) alike have challenged this view. Most recently, Carruthers (2011) has put forward a detailed empirically oriented argument against it.

He holds that if subjects had direct self-knowledge of attitudes then there shouldn't be cases where they unknowingly confabulate attitudes for their own behaviour. But he points out that various psychological studies show that there are. For instance, in experimental settings, subjects unconsciously confabulated decisions for their own actions after the latter occurred (Gazzaniga, 1995; Johansson *et al.*, 2006) and even though they didn't perform the actions themselves (Wegner and Wheatley, 1999) or voluntarily (Brasil-Neto *et al.*, 1992).

Furthermore, Carruthers continues, if subjects knew their own attitudes directly then self-interpretation shouldn't unbeknownst to them affect their self-ascriptions of attitudes. But he holds that again a number of studies show that it does. It has been found, for instance, that interpretation of one's own body posture (Briñol *et al.*, 2009) and body movements (e.g. head-nodding) affects self-ascriptions of attitudes (Briñol and Petty, 2003).

[2] In Section 2 I will say more on what is meant by 'conscious' in this paper.

Carruthers argues that given these (and many more) findings on confabulation of attitudes and self-interpretation, the intuition that we have non-interpretive self-knowledge of attitudes is undermined. The reason is that the subjects in the mentioned studies were unaware of confabulating attitudes and interpreting themselves. From the first-person point of view, it can thus *seem* as if one knows one's own attitudes without interpreting oneself even when self-interpretation or confabulation of attitudes occurs. The intuition that one has direct self-knowledge of attitudes is hence in need for support.

Carruthers holds that the empirical data and various theoretical considerations (drawn, for example, from the global workspace theory, accounts of sensory-based working memory, and evolutionary theories on meta-representation, see Carruthers, 2011, pp. 48ff.) support an alternative view, what he calls an 'interpretive sensory access' or 'ISA' theory of self-knowledge. According to the ISA theory, there is no principled difference between self-knowledge of attitudes and knowledge of other people's mental states. Rather, in both cases, knowledge of attitudes is interpretive in nature and the result of the operation of the same cognitive system, the mindreading faculty, which, based on observation and interpretation of a subject, issues judgments about that subject's mental states. Unlike in the case of knowledge of others' mental states, however, in one's own case, the mindreading faculty can access more information for its interpretation. In addition to overt behaviour and a subject's circumstances, it can also utilize a subject's affective, sensory and, in particular, imagistic states (e.g. visual imagery, or 'inner speech').

For the purpose of this paper, I shall assume that the ISA theory is correct. What I want to focus on is Carruthers' case against conscious attitudes.

2. Theories of Consciousness and the Case Against Conscious Attitudes

Carruthers holds that 'there are two broad possibilities for accounting for the conscious status of propositional attitudes': these are either the 'higher-order state approach' or the 'global workspace theory' (Carruthers, 2011, p. 374). He claims that given this 'disjunction of possible accounts of attitude consciousness...., the ISA theory entails' that there are no conscious attitudes³ (*ibid.*, p. 378).

[3] Recall that the exceptions are 'sensorily-embedded judgments' and 'context-bound desires and emotions. These are globally broadcast and transparently accessible to the mindreading system' (Carruthers, 2011, p. 378).

Before getting into the details of Carruthers' argument a clarification is in order, for a state's being conscious can mean at least two different things. It can mean the state's being *phenomenally* conscious, which involves its having experiential properties or a 'what-it's-likeness' (e.g. it is like something to smell fresh coffee, feel pain, etc.); or it can mean the state's being *access* conscious, which needn't involve its having experiential properties but only requires that the state be broadcast for reasoning, action planning, and verbal report (see Block, 2002).

Both the higher-order state approach and the global workspace theory are often taken to be accounts of phenomenally conscious states. But some theorists might reject this view and hold that, for instance, the global workspace theory accounts at best for access conscious states.

In what follows nothing hinges on this issue, for whatever kind of consciousness one takes the two approaches to explain, Carruthers' claim is that these approaches preclude the existence of conscious attitudes if attitudes are only known interpretively. I shall show that, no matter what kind of consciousness is at issue, this is not the case. I begin by introducing and critiquing his argument with respect to the higher-order state approach before turning to the global workspace theory.

2.1. *The higher-order state approach*

According to the higher-order state approach to consciousness, a mental state *M* is conscious in virtue of another state that is about *M*. There are different higher-order state views depending on how the higher-order state at issue is specified. For instance, some philosophers hold that it is an inner-sense perception of *M* (e.g. Armstrong, 1981; Lycan, 1996) while others argue that it is a thought about *M* (e.g. Carruthers, 2000; Rosenthal, 1997; 2005). Nonetheless, the different higher-order theories have in common that *M* is conscious only if the subject, who is in *M*, is aware of *M*.

Furthermore, Carruthers holds that '[m]ost higher-order theories of phenomenal consciousness entail' that the 'access to our own experiences' that is involved in this awareness 'is transparent, and radically different from the sort of interpretive access that we have to the experiences of other people. Indeed, this is believed by most people to be an important mark in favor of the approach' (Carruthers, 2011, p. 376).

Carruthers notes that the only exception is David Rosenthal's (2005) higher-order thought account. According to Rosenthal, as long as the way in which a higher-order thought about *M* is produced doesn't involve any *conscious* inferences or interpretations, this will still be sufficient to make *M* conscious.

But Carruthers responds that this view implies that 'it is sufficient for one to be undergoing a phenomenally conscious pain, say, that one should come to *believe* that one is in pain on the basis of unconscious inferences grounded in observation of one's own circumstances and behavior' (Carruthers, 2011, p. 376). And this, he holds, is 'extremely hard to accept' (*ibid.*). For the purposes of his argument, he thus assumes that

if propositional attitudes are to count as conscious, according to a higher-order account, then our access to those attitudes isn't interpretive, and must occur independently of beliefs about our own circumstances, behavior, and other mental states. (*Ibid.*, p. 376)

Carruthers continues that if 'conscious attitudes would have to be attitudes that one knows of *without* relying on self-interpretation',⁴ then since the evidence suggests and the ISA theory of self-knowledge implies 'that the only way in which one can know' of one's own attitudes 'is by mindreading inferences', i.e. by self-interpretation, it follows that there are no conscious attitudes on the higher-order state approach (*ibid.*, p. 378).

As it stands, however, this seems to conflate access to one's own attitudes with self-knowledge of them. As just mentioned, Carruthers holds that, on the higher-order state account, attitudes that are the targets of a higher-order state can only be conscious if 'our *access* to those attitudes isn't interpretive' (*ibid.*, p. 376, emphasis added). He then points out that, given the empirical evidence and theoretical support for the ISA theory, 'the only way in which one can *know*' (*ibid.*, p. 378, emphasis added) of one's own attitudes is by self-interpretation and concludes from this that there are no conscious attitudes on the higher-order account. Unless access to one's own attitudes is equated with self-knowledge of them, however, this doesn't follow.

If the two were identical and could be equated then there would be no problem for Carruthers' argument. But they need to be kept separate. To begin with, self-knowledge of a belief that *p* requires a conceptualization of the belief and a self-ascription with the content 'I believe that *p*'. Furthermore, self-knowledge and knowledge more generally is, as Goldman (2006, p. 224) puts it, a 'thick epistemo-

[4] If not otherwise indicated, italics in the quotes of this paper are original.

logical concept' which refers to more than just attribution or belief but also to justified true belief, or reliably formed true belief. In contrast, *access* to the belief that *p* doesn't require any of this. For instance, one might have access to one's own attitude without conceptualizing and self-ascribing the attitude as such or at all. One way to specify the proposal is by holding that one counts as having access to one's own attitudes when they are broadcast in the workspace; that is a first-order view I discuss below.

Here are two higher-order state accounts of attitude consciousness that specify this further.

One might hold that the higher-order thought that makes a first-order attitude conscious doesn't involve a conceptualization of the attitudes as an attitude but rather takes the form 'I am in *that_*' — where '*that_*' is a mental demonstrative picking out a particular first-order attitude without conceptualizing it as such.⁵ Since self-knowledge of attitudes requires conceptualization of attitudes as attitudes, and since a demonstrative higher-order thought doesn't require that, even if, as Carruthers holds, self-knowledge of attitudes is interpretive in nature, it doesn't follow that the formation of the demonstrative higher-order thought requires self-interpretation also. Thus, there is a higher-order thought account of attitude consciousness that is unaffected by Carruthers' argument.⁶

Furthermore, higher-order perception views of attitude consciousness (e.g. Armstrong, 1981; Lycan, 1996) are not threatened by his argument either. For perceptual states needn't involve a conceptualization of what is perceived. Similarly, a higher-order perception of an attitude that makes the latter conscious needn't amount to a conceptualization of the attitude. That is, there could be higher-order awareness of an attitude without self-ascription and knowledge of that attitude. Since that is so, and since the ISA theory only pertains to self-ascriptions (i.e. to awareness involving a conceptualization of an attitude as a particular attitude), Carruthers' argument against conscious attitudes doesn't succeed.

It may be objected that the preceding discussion overlooks Carruthers' (2011, pp. 64–8) point that there is no independent ground for assuming that there exists in addition to the mindreading faculty a second mechanism that, unlike the mindreading faculty, produces

-
- [5] Lurz (2006) offers an account of conscious attitudes along similar lines but first-order in nature.
- [6] Some higher-order thought theories might require that a conscious attitude be an attitude that is known to the subject non-interpretively. With respect to these higher-order thought theories, Carruthers' argument would still hold.

higher-order states that don't involve an attitude conceptualization. A possible response here is that such a mechanism is more fundamental and required to cause contents to be broadcast and widely available to systems such as the mindreading faculty. I shall say more on this in the next section.

For now, I conclude that once access to and self-knowledge of attitudes are kept separate, Carruthers' general claim that if the higher-order state approach is combined with the ISA theory then this conjunction of theories yields the result that there are no conscious attitudes is false. There are higher-order state approaches that allow for conscious attitudes even if the latter are only known interpretively.

2.2. *Global workspace theory*

But Carruthers' point might still go through with respect to the second account of consciousness that he considers, the global workspace theory.

According to the global workspace theory (Baars, 1988; Dehaene and Naccache, 2001), the mind is made up of specialist sensory and conceptual systems that are connected to a 'workspace' in the prefrontal cortex of the brain. This is an area across which the systems can broadcast their outputs. Any content that is broadcast in the workspace is automatically accessible to and usable by all systems that are connected to the workspace. It is via the workspace that memory, motivational, and judgment and decision making systems exchange contents and communicate.

The global workspace theory has been proposed as an account of cognition as well as consciousness. On this view, a mental state is *conscious* if it is in the workspace. As Dehaene and Changeux (2011, p. 210) put it, the 'global availability of information... is what we subjectively experience as a conscious state'. And a mental state counts as being in the workspace, if it is widely available to processes involved in, for example, reasoning and decision making.

Given this, Carruthers argues that if there were conscious attitudes on the global workspace theory then these attitudes would also have to be states that are in the workspace and hence 'globally accessible to all of the main executive processes of the mind' (Carruthers, 2011, p. 375). But, he continues, no attitudes 'are ever globally accessible in the mind-brain. For if they were, then of course they would be available as input to the mindreading faculty among other systems, and they would thus be accessible to be known independently of any associated sensory representations', i.e. subjects should be able to know

them directly, non-interpretively (*ibid.*, p. 378). But the evidence suggests and the ISA theory implies that attitudes are not non-interpretively knowable. Thus, Carruthers concludes, there are no attitudes in the global workspace, and there are no conscious attitudes according to the global workspace theory (*ibid.*).

However, even if we accept that the mindreading faculty lacks direct access to attitudes, it doesn't follow that the latter aren't in the workspace. The additional assumption required here is that the mindreading faculty has direct access to *all* the states in the global workspace. And this assumption can be challenged. One could hold, for instance, that the faculty has access only to a particular subset of all the states in the workspace, where attitudes are not in that set. One of Carruthers' (*ibid.*, pp. 64–8) own considerations helps motivate this view. He argues, for example, that the mindreading faculty evolved specifically for the purpose of interpretively working out mental states and only consumes contents required for that function — which includes visual imagery, inner speech, perceptions of behaviour, etc. but excludes a subject's own attitudes.⁷ Given this, Carruthers' inference from the faculty's indirect access to attitudes to the claim that there are no attitudes in the workspace can be rejected.

He does, however, offer other reasons against globally broadcast attitudes that don't directly pertain to self-knowledge of attitudes. He argues, for instance, that all 'of the evidence that has been accumulated in support of global broadcast theory... concerns the global broadcast of sensory information in the brain (together with conceptual information that is bound into the contents of sensory states, of course)',⁸ and no 'evidence has been presented' that attitudes, too, 'can be globally broadcast, except by first being formulated into a sensory image of some sort, such as a sentence of inner speech' (Carruthers, 2011, p. 54).

One way to respond to Carruthers here and to argue for globally broadcast attitudes involves appealing to the fact that at least some attitudes are 'inferentially promiscuous', i.e. able to enter into various inferential relations and interact with one another (Stich, 1978; Fodor, 2000; Evans, 1982; Brewer, 1999; Hurley, 2006). Since the inferential promiscuity of attitudes requires them to be available to each other, this property of them suggests that they can be broadcast in a workspace.

[7] Note that it does include mindreading domain-specific attitudes (Carruthers, 2011, pp. 53–4).

[8] See Carruthers (2011, p. 48) for details on what is meant by 'bound into'.

Carruthers anticipates this point. In reply, he holds that the inferential promiscuity of attitudes can be explained without assuming that they themselves are able to enter the workspace:

[N]o propositional attitudes figure in the central workspace. Their interactions with other such attitudes are always indirect, mediated by processes that create sensory imagery of various kinds that *can* enter the global workspace. (Carruthers, 2013, p. 3)

But how exactly are attitudes supposed to interact by means of imagistic states? Carruthers insists that no such state can itself be an attitude, for any imagistic state will lack the right functional profile to be an attitude (Carruthers, 2011, pp. 102ff.). To illustrate, suppose upon evaluating various hypotheses about the weather, you come to entertain visual imagery of a rainy day and rehearse in inner speech the utterance ‘It will rain today’. Carruthers concedes that these imagistic states might resemble a judgment, but he claims that no such states are judgments themselves. For judgments proper terminate theoretical reasoning on an issue and are directly available for action, yet this doesn’t hold for any kind of imagery. You might, for instance, say to yourself in inner speech ‘It will rain today’ even though you don’t judge this to be so and won’t act in accordance with such a judgment. Carruthers holds that further reasoning is required for imagistic states to settle an issue and lead to action. In his view, what gives an imagistic state, for example, ‘a belief-like’ as opposed to, say, ‘a supposition-like causal role will depend on one’s interpretation of its nature’, i.e. it will depend on mindreading (Carruthers, 2013, p. 10; see also Carruthers, 2011, pp. 102–17). Given this, attitude interaction by means of sensory-imagery will then also require mindreading, for ‘sensory representations in general need to be interpreted in order to be classified as involving one sort of mental attitude rather than another’ (Fletcher and Carruthers, 2012, p. 13740).

3. Indirect Attitude Interaction and Inferential Reasoning in Autism

Is Carruthers’ account of attitude interaction plausible? As it turns out, there is evidence that speaks against his proposal and provides the basis for a positive argument for conscious attitudes.

Before going into the details, however, note that, as it stands, Carruthers’ claim that interactions of ‘attitudes are *always* indirect’ and proceed by means of imagistic states (Carruthers, 2013, p. 3, emphasis added) can’t be right if such states ‘in general need to be interpreted’ (Fletcher and Carruthers, 2012, p. 13740). For the result

of an interpretation of imagery is itself a judgment: a judgment produced by the mindreading faculty concerning the attitude underlying the imagery. Since judgments are themselves attitudes, this judgment could then, if the above general claim were correct, again only interact with other attitudes by means of imagistic states (which again would have to be interpreted, resulting in a further judgment, etc.). Attitude interaction would become impossible. Since attitude interaction isn't impossible, it follows that some attitudes must be able to enter into inferences and interact without first becoming expressed in interpretation-dependent imagery.

In other places (see Carruthers, 2011, pp. 53–4, 71), Carruthers in fact grants the point that some attitudes can interact directly, and restricts his claim only to attitudes that are *widely available* to judgment and decision-forming systems. It is these attitudes in particular that are only thus available in virtue of interpretation-dependent imagery being broadcast in the workspace, he holds.

To assess this proposal, we first need to find a task that involves widely available attitudes. Consider the following. Suppose you are asked whether Jack owns more money than James, and engage in explicit reasoning on the matter. Suppose that you know that (i) Jack has more money than Jill, and also recall that a reliable friend told you that (ii) Jill owns more money than James. Suppose you then infer that Jack owns more money than James, give an affirmative answer to the question, and mention the beliefs (i) and (ii) as reasons for your answer.

To answer the question, you have to draw a transitive inference from the beliefs (i) and (ii). These beliefs thus interact with each other. But are they plausibly viewed as being widely available?

It might be pointed out that animals such as rats, pigeons, and primates are able to perform transitive inferences too (Davis, 1992; von Fersen *et al.*, 1991; Wynne, 1995), and that this suggests that these inferences don't require cognitive sophistication. Perhaps they don't depend on widely available attitudes or the broadcast of them in the workspace.

However, note that in the example above you also invoke the beliefs to justify your judgment on and answer to the question concerning Jack and James. Since they thus enter into personal-level processing and underlie report, there is good ground to hold that the beliefs *are* widely available and in the workspace.

On Carruthers' view, however, only sensory-imagistic states (and conceptual contents embedded in them)⁹ are in the workspace and attitudes can only interact by imagistic states that need to be interpreted by the mindreading faculty. If Carruthers is right then subjects who have an impaired mindreading faculty and fail at ascribing attitudes should not exhibit normal performance in tasks that require reasoning along the lines just mentioned.

This prediction turns out to be false, however. Consider autism. A number of studies have shown that people with autism tend to fail at ascribing beliefs¹⁰ to themselves and others (Baron-Cohen, 1995; Frith and Happé, 1999; Williams, 2010). Yet, there is evidence that they perform normally in tasks that require various kinds of first-order reasoning (see, for example, Baron-Cohen *et al.*, 2001; Scott *et al.*, 1996; 1999). For instance, Scott *et al.* (1996) gave autistic and a control group of children two tests. The first one required abstract reasoning involving transitive inferences (i.e. $X \rightarrow Y$, $Y \rightarrow Z$; hence $X \rightarrow Z$) and analogical reasoning (e.g. X is to Y as P is to Q, etc.). The second test that Scott *et al.* gave the children was a standard mindreading task (the Sally-Anne test, see Baron-Cohen *et al.*, 1985; Wimmer and Perner, 1983) in which subjects had to attribute a false belief to another agent. As it turned out, most autistic children failed in the mindreading task. This suggests that they lacked an understanding of belief as a representational state underlying a subject's action. Such understanding is required for correct other- as well as self-ascriptions of attitudes. Interestingly, however, they 'performed comparably to the control groups, both on a test of transitive inferential reasoning and on a test of analogical reasoning' (Scott *et al.*, 1996, p. 235). Note that after each answer to one of the questions in the reasoning test, Scott *et al.* also asked the children for a justification for their answer. This was to ensure that they were not merely guessing but actually engaging in inferential reasoning. Scott *et al.* found no significant difference between the justifications that the autistic children and the controls provided for their answers.

Since the reasoning that Scott *et al.* tested required making connections between several beliefs, and since the autistic subjects were able to justify their subsequent judgments by mentioning the contents of

[9] In what follows I shall use the term 'sensory-imagistic state' to include both kinds of content, i.e. imagistic contents, which might be non-conceptual, and conceptual information that is bound into the contents of imagistic states.

[10] This is not to say that all of them do. High-functioning autistic individuals might be able to ascribe beliefs correctly. But note that even they appear to employ a different strategy to do so than normal subjects do (Frith and Happé, 1999).

these beliefs, the results provide good ground to believe that the latter interacted with each other and were widely available.¹¹ Furthermore, since these subjects exhibited a lack of understanding of beliefs and hence a defunct mindreading faculty, Scott *et al.*'s findings speak against the view that widely available attitudes can only interact by imagistic states that are interpretation- (and hence mindreading-) dependent. That is, Scott *et al.*'s results speak against Carruthers' view of attitude interaction.¹²

Could it be that the attitudes at issue interacted by imagistic states that didn't have to be interpreted? There are two points to note about this suggestion. First, if attitudes could interact by imagistic states that needn't be interpreted for their underlying attitudes then since the mindreading faculty has non-interpretive access to one's own imagery (Carruthers, 2011, p. 2), these imagistic states should also give the faculty non-interpretive access to attitudes by means of these states. Carruthers wishes to deny non-interpretive access to attitudes, however. The suggestion under consideration is thus not available to him. Second, if attitudes interacted by imagistic states that needn't be interpreted then these states would lead directly to inferential transitions, judgment- and decision-formation, etc. just as beliefs and other attitudes do. That is, they would at the personal level play the same functional role as attitudes and there would be little reason to deny that they qualify as attitudes themselves.¹³ Note that if (occurrent) attitudes are imagistic states, they could be in the workspace even if one held that on the currently best-supported account of global workspace only sensory-imagistic states are broadcast in the workspace.

In sum, then, findings on reasoning in autism suggest that attitudes are at least sometimes directly available to various judgment- and decision-forming systems and broadcast in the workspace even when a subject lacks an understanding of attitudes that is required for self-ascriptions and self-knowledge of them. If we assume the global

[11] It might be objected that if autistic subjects did indeed have widely available attitudes then they should exhibit at least normal performance on executive function tasks, i.e. on tasks that require inhibiting responses and updating beliefs/desires in the light of new information, but in a number of studies it has been shown that autistic subjects have clear executive function deficits (see, for example, Gioia *et al.*, 2002; Geurts *et al.*, 2009). However, note that flexibility in thought, i.e. revising beliefs with new evidence, is one thing; their being widely available for reasoning, verbal report, etc. is another. For all we know, in autistic subjects, the mechanism responsible for revising beliefs in light of new information might be dysfunctional even though their beliefs are still widely available and used in first-order reasoning tasks.

[12] The autism data also challenge Frankish's (2012) dual-attitudes account, for the latter involves a commitment to higher-order beliefs (see Frankish, 2012, p. 47).

[13] But see Frankish (2012) and Carruthers (2013) for further discussion on this point.

workspace theory of consciousness, this supports the view that there are conscious attitudes¹⁴ even in cases when a subject doesn't have an understanding of them. The view that subjects come to know their own attitudes only by self-interpretation (or not at all) thus doesn't undermine the existence of conscious attitudes on the global workspace view.

4. Conscious Attitudes and Logical Reasoning

There is a second set of studies that further support the view that there are conscious attitudes. The studies concern logical reasoning under cognitive load in normal subjects.

To a first approximation, suppose that only non-attitude states (e.g. sensory-imagistic states) are conscious and that attitudes are always unconscious. If this were so then keeping a subject's unconscious thoughts occupied during a task that requires attitude interaction should negatively affect the performance in that task. Personal-level logical reasoning that is involved in solving symbolic logic puzzles, for instance, requires moving from suppositions and beliefs to judgments. It thus involves attitude interaction and should be negatively affected by the mentioned manipulation.

There is evidence against this prediction, however. DeWall *et al.* (2008) conducted a study that required subjects to solve logic puzzles under different cognitive load conditions. In one condition, DeWall *et al.* impeded conscious processing by asking subjects to solve the puzzles and simultaneously count the word 'time' in a song that they were listening to during the task. As it turned out, the subjects performed significantly worse than no-load controls. This suggests that conscious processing is required for logical reasoning. In a second and here more relevant condition, DeWall *et al.* then manipulated non-conscious cognition during the reasoning task. The procedure they used was based on Wegner's (1994) ironic processing theory, which concerns the cognitive process that allows subjects to deal with unwanted thought contents by suppressing them. According to the theory, thought suppression has two components. It involves a non-conscious monitor that keeps track of cues that might evoke unwanted thoughts, and a conscious suppression mechanism. Crucially, Wegner found that the non-conscious monitor keeps scanning the subject's thoughts for unwanted cues even after the subject's conscious attention is turned elsewhere. In some cases, when the resources of the

[14] As noted above, for present purposes, *access* conscious attitudes suffice.

conscious suppression mechanism are depleted, this has the ironic effect that the unwanted thought contents occur in the subject's consciousness more frequently than before (*ibid.*).

Building on Wegner's theory, DeWall *et al.* used thought suppression to manipulate unconscious processing. They had their test subjects think of an 'old flame', a former relationship partner, and then instructed them to stop thinking about him or her. This freed up the subjects' conscious processing so that they could solve logic problems while the unconscious monitoring system was still engaged with thinking about the old flame. That the unconscious system was indeed occupied throughout was verified by a post-task measure that showed that the thought of the old flame remained highly accessible after the logic problems.

The results of the experiment were that subjects whose non-conscious processing was impeded with suppressed thoughts about an old flame didn't perform worse than no-load control subjects. In fact, the non-conscious load subjects solved more logic problems correctly than subjects in the no-load control condition. These findings contradict the prediction that non-conscious load would impede logical reasoning.

But even if the thought suppression task didn't negatively affect logical reasoning, it might still be that logical reasoning was performed by an unconscious system that is isolated from the processing involved in unconscious thought monitoring.

However, if there were an unconscious system performing logical reasoning then we would expect that priming subjects with logic-specific terms should positively affect performance in the reasoning task, for this should activate and prepare the system before the task. DeWall *et al.* (2008) tested this further prediction. They found that, while non-conscious activation of the idea of logical reasoning did increase the activation of logic-relevant concepts, it did not improve the subjects' performance. In fact, DeWall *et al.* noticed a trend that subjects who were not logic primed were doing better on the reasoning task than those that were primed.

Taken together, the findings just reviewed support the following indirect argument for conscious attitudes. If there were no conscious attitudes then attitudes would have to interact unconsciously. If they could only interact unconsciously, then, since logical reasoning involves attitude interaction, logical reasoning would involve unconscious attitude interaction. Given this, impeding unconscious thinking in a logical reasoning task should negatively affect task performance, and priming the subject for logical reasoning should positively affect

it. However, the results of DeWall *et al.*'s study speak against both of these predictions and thus suggest that at least some attitudes (those involved in logical reasoning) are conscious.

5. Conclusion

Carruthers holds that the conjunction of (i) the view that self-knowledge of attitudes is interpretive, and (ii) either of the 'two broad possibilities for accounting' for conscious attitudes (i.e. the higher-order state approach and the global workspaces theory) implies that there are no conscious attitudes (Carruthers, 2011, p. 374). I showed that this is not the case, for the two accounts of consciousness don't require that a conscious attitude be one that is non-interpretively known to the subject. I argued furthermore that studies on reasoning in autistic and control subjects provide positive grounds for believing that there are conscious attitudes.¹⁵

References

- Armstrong, D. (1981) *The Nature of Mind and Other Essays*, Ithaca, NY: Cornell University Press.
- Baars, B. (1988) *A Cognitive Theory of Consciousness*, New York: Cambridge University Press.
- Baron-Cohen, S. (1995) *Mindblindness*, Cambridge, MA: MIT Press.
- Baron-Cohen, S., *et al.* (1985) Does the autistic child have a 'theory of mind'?, *Cognition*,
- Baron-Cohen, S., *et al.* (2001) Are intuitive physics and intuitive psychology independent? A test with children with Asperger Syndrome, *Journal of Developmental and Learning Disorders*, **5**, pp. 47–78.
- Bem, D. (1967) Self-perception: An alternative interpretation of cognitive dissonance phenomena, *Psychological Review*, **74**, pp. 183–200.
- Block, N. (2002) Some concepts of consciousness, in Chalmers, D.J. (ed.) *Philosophy of Mind*, Oxford: Oxford University Press.
- Brasil-Neto, J., *et al.* (1992) Focal transcranial magnetic stimulation and response bias in a forced choice task, *Journal of Neurology, Neurosurgery, and Psychiatry*, **55**, pp. 964–966.
- Brewer, B. (1999) *Perception and Reason*, Oxford: Oxford University Press.
- Briñol, P. & Petty, R. (2003) Overt head movements and persuasion: A self-validation analysis, *Journal of Personality and Social Psychology*, **84**, pp. 1123–1139.
- Briñol, P., *et al.* (2009) Body posture effects on self-evaluation: A self-validation approach, *European Journal of Social Psychology*, **39**, pp. 1053–1064.
- Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge: Cambridge University Press.
- Carruthers, P. (2011) *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford: Oxford University Press.

[15] Parts of this paper were presented as a poster at the Ruhr University Bochum and at the ESPP meeting in Granada in 2013. I would like to thank the audiences on these occasions for helpful discussions. Many thanks also to two referees for this journal for useful comments.

- Carruthers, P. (2013) On central cognition, *Philosophical Studies*, **160**, pp. 1–20.
- Davis, H. (1992) Transitive inference in rats (*Rattus norvegicus*), *Journal of Comparative Psychology*, **106**, pp. 342–349.
- Dehaene, S. & Naccache, L. (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework, *Cognition*, **79**, pp. 1–37.
- Dehaene, S. & Changeux, J. (2011) Experimental and theoretical approaches to conscious processing, *Neuron*, **70**, pp. 200–227.
- Dennett, D. (1991) *Consciousness Explained*, Boston, MA: Little Brown.
- DeWall, C., et al. (2008) Evidence that logical reasoning depends on conscious processing, *Consciousness and Cognition*, **17**, pp. 628–645.
- Evans, G. (1982) *The Varieties of Reference*, Oxford: Oxford University Press.
- Fletcher, L. & Carruthers, P. (2012) Metacognition and reasoning, *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, pp. 1366–1378.
- Fodor, J. (2000) *The Mind Doesn't Work That Way*, Cambridge, MA: MIT Press.
- Frankish, K. (2012) Dual systems and dual attitudes, *Mind and Society*, **11**, pp. 41–51.
- Frith, U. & Happé, F. (1999) Theory of mind and self consciousness: What is it like to be autistic?, *Mind & Language*, **14**, pp. 1–22.
- Gazzaniga, M. (1995) Consciousness and the cerebral hemispheres, in Gazzaniga, M. (ed.) *The Cognitive Neurosciences*, Cambridge, MA: MIT Press.
- Geurts, H., et al. (2009) The paradox of cognitive flexibility in autism, *Trends in Cognitive Sciences*, **13**, pp. 74–82.
- Gioia, G., et al. (2002) Profiles of everyday executive function in acquired and development disorders, *Child Neuropsychology*, **8**, pp. 121–137.
- Goldman, A. (2006) *Simulating Minds*, Oxford: Oxford University Press.
- Gopnik, A. (1993) How we know our own minds: The illusion of first-person knowledge of intentionality, *Behavioral and Brain Science*, **16**, pp. 1–14.
- Gopnik, A. & Meltzoff, A. (1994) Minds, bodies, and persons: Young children's understanding of the self and others as reflected in imitation and theory of mind research, in Parker, S., Mitchell, R. & Boccia, M. (eds.) *Self-awareness in Animals and Humans*, New York: Cambridge University Press.
- Hurley, S. (2006) Making sense of animals, in Hurley, S. & Nudds, M. (eds.) *Rational Animals?* Oxford: Oxford University Press.
- Johansson, P., et al. (2006) How something can be said about telling more than we can know: On choice blindness and introspection, *Consciousness and Cognition*, **15**, pp. 673–692.
- Lawlor, K. (2008) Self-knowledge of belief: Seeking causes, in Brakel, L. (ed.) *American Imago*, **65**, pp. 335–356.
- Lawlor, K. (2009) Knowing what one wants, *Philosophy and Phenomenological Research*, **79**, pp. 47–75.
- Lurz, R. (2006) Conscious beliefs and desires: A same-order approach, in Kriegel, U. & Williford, K. (eds.) *Self-Representational Approaches to Consciousness*, Cambridge, MA: MIT Press.
- Lycan, W. (1996) *Consciousness and Experience*, Cambridge, MA: Bradford Books/MIT Press.
- Nichols, S. & Stich, S. (2003) *Mindreading*, Oxford: Oxford University Press.
- Rosenthal, D. (1997) A theory of consciousness, in Block, N., Flanagan, O. & Güzeldere, G. (eds.) *The Nature of Consciousness*, Cambridge, MA: Bradford Books/MIT Press.
- Rosenthal, D. (2005) *Consciousness and Mind*, Oxford: Oxford University Press.
- Ryle, G. (1948) *The Concept of Mind*, New York: Barnes & Noble.

- Scott, F., *et al.* (1996) Logical, analogical, and psychological reasoning in autism: A test of the Cosmides theory, *Development and Psychopathology*, **8**, pp. 235–246.
- Scott, F., *et al.* (1999) ‘If pigs could fly’: A test of counterfactual reasoning and pretence in children with autism, *British Journal of Developmental Psychology*, **17**, pp. 349–362.
- Shoemaker, S. (1996) *The First-Person Perspective and Other Essays*, Cambridge, MA: Cambridge University Press.
- Stich, S. (1978) Beliefs and subdoxastic states, *Philosophy of Science*, **45**, pp. 499–518.
- von Fersen, L., *et al.* (1991) Transitive inference in pigeons, *Journal of Experimental Psychology: Animal Behavior Processes*, **17**, pp. 334–341.
- Wegner, D. (1994) Ironic processes of mental control, *Psychological Review*, **101**, pp. 34–52.
- Wegner, D. (2002) *The Illusion of Conscious Will*, Cambridge, MA: Harvard University Press.
- Wegner, D. & Wheatley, T. (1999) Apparent mental causation: Sources of the experience of will, *American Psychologist*, **54**, pp. 480–491.
- Williams, D.M. (2010) Theory of own mind in autism: Evidence of a specific deficit in self-awareness?, *Autism: International Journal of Research and Practice*, **14**, pp. 474–494.
- Wilson, T.D. (2002) *Strangers to Ourselves*, Cambridge, MA: Harvard University Press.
- Wimmer, H. & Perner, J. (1983) Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception, *Cognition*, **13**, pp. 103–128.
- Wynne, C. (1995) Reinforcement accounts for transitive inference performance, *Animal Learning & Behavior*, **23**, pp. 207–217.

Paper received August 2012; revised December 2013.