



Interpretive sensory-access theory and conscious intentions

Uwe Peters

To cite this article: Uwe Peters (2014) Interpretive sensory-access theory and conscious intentions, *Philosophical Psychology*, 27:4, 583-595, DOI: [10.1080/09515089.2012.749560](https://doi.org/10.1080/09515089.2012.749560)

To link to this article: <https://doi.org/10.1080/09515089.2012.749560>



Published online: 24 Jan 2013.



Submit your article to this journal [↗](#)



Article views: 194



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)

Review Essay

Interpretive sensory-access theory and conscious intentions

The Opacity of Mind: An Integrative Theory of Self-Knowledge

Peter Carruthers

Oxford: Oxford University Press, 2011

448 pages, ISBN: 0 199 596 190 (hbk); \$55.00

Uwe Peters

*It is typically assumed that while we know other people's mental states by observing and interpreting their behavior, we know our own mental states by introspection, i.e., without interpreting ourselves. In his latest book, *The opacity of mind: An integrative theory of self-knowledge*, Peter Carruthers [2011. *The opacity of mind: An integrative theory of self-knowledge*. Oxford: Oxford University Press] argues against this assumption. He holds that findings from across the cognitive sciences strongly suggest that self-knowledge of conscious propositional attitudes such as intentions, judgments, and decisions involves a swift and unconscious process of self-interpretation that utilizes the same sensory channels that we employ when working out other people's mental states. I provide an overview of Carruthers' book before discussing a pathological case that challenges his account of self-knowledge and mentioning empirical evidence that undermines his use of a particular kind of data in his case against introspection of conscious attitudes.*

Keywords: Conscious Attitudes; Introspection; Self-Knowledge

1. Introduction

In folk psychology, as in the cognitive sciences and philosophy, it is commonly assumed that while we know other people's mental states by observing and interpreting their behavior, we know our own mental states by introspection, that is, without interpreting ourselves (Farkas, 2008; Goldman, 2006; Nichols & Stich, 2003; Searle, 1983; Shoemaker, 1996). Even though the assumption of introspective

Uwe Peters is a PhD student at King's College London.

Correspondence to: Uwe Peters, Department of Philosophy, King's College London, Strand, London WC2R 2LS, UK. Email: uwe.peters@kcl.ac.uk

This article was originally published with errors. This version has been amended. Please see Corrigendum (10.1080/09515089.2013.848049).

self-knowledge is prevalent, it is not without its critics (Dennett, 1992; Hirstein, 2005; Ryle, 1949; Stephens & Graham, 2000).

The latest attack on introspection can be found in Peter Carruthers' (2011) book *The opacity of mind: An integrative theory of self-knowledge*. Carruthers grants that we have non-interpretive access to our own sensory and affective states,¹ but he denies that this is the case with respect to propositional attitudes (PAs)² such as judgments, intentions, and decisions. Carruthers holds that self-knowledge of PAs involves a swift and unconscious process of self-interpretation that makes use of the same sensory channels that we utilize when working out other people's mental states.

In the following, I briefly summarize Carruthers' case against introspection and for his own account of self-knowledge of PAs (section 2) before mentioning two critical comments. I argue that there is a pathological case that challenges Carruthers' account of self-knowledge (section 3). Furthermore, there is empirical evidence that undermines his use of a particular kind of data in his case against introspection of PAs (section 4).

2. Interpretive Sensory-Access Theory and the Argument Against Introspection

Carruthers' main argument in *The opacity of mind* takes the form of an inference to the best explanation. He contends that a wealth of cognitive scientific evidence speaks strongly against introspective access to one's own PAs and is most adequately explained by his "indirect sensory access" (ISA) theory, according to which self-knowledge of PAs involves turning the "mindreading faculty"—the faculty that allows us to attribute mental states to others and predict their behavior in terms of them—towards oneself. Carruthers holds (chapter 1) that the only difference between self- and other-knowledge of PAs is that in one's own case, the mindreading faculty has more information available upon which to base its interpretation. In addition to using overt behavior, in one's own case, it can also draw on a subject's affective, sensory, and quasi-sensory states such as visual imagery or "inner speech" tokens that are globally broadcast in the mind-brain.

But why assume that the assumption of introspective access to PAs is mistaken? Carruthers begins his case against introspection and in support of his ISA theory by discussing findings from Gazzaniga's (1995) split-brain³ experiments. In one of the experiments, different stimuli were presented to the two brain hemispheres of a split-brain patient, "Joe," at the same time. When the instruction "walk!" was flashed to Joe's right hemisphere, he got up and walked out of the van in which the experiment took place. When asked why he left the van, Joe swiftly and sincerely replied that he wanted to go to the house to get a Coke, even though his behavior was in fact triggered by the instruction to walk that was flashed to his right brain hemisphere. His intention was, unbeknownst to him, confabulated.

Carruthers contends that for all we know, we, like Joe, may merely have the impression that we introspect our own PAs such as intentions while we are, in fact, unable to do so. In the absence of evidence for introspection of PAs, the findings

from Gazzaniga's split-brain experiments undermine the intuition of introspection, Carruthers maintains. He holds that unless we are given convincing reasons for believing in introspection, this intuition should no longer be accepted as true, and should not bias our evaluation of competing theories of self-knowledge.

In support of his ISA theory of self-knowledge, Carruthers then argues (chapter 2) that his account is not only well in line with other theories in the cognitive sciences such as global workspace accounts of our cognitive architecture (Baars, 1988; Dehaene & Naccache, 2001), models of working memory (Baddeley & Logie, 1999; Gathercole & Baddeley, 1993) and accounts of the evolution of the mindreading faculty (Byrne & Whiten, 1997). Furthermore, he holds that since it doesn't require two separate mechanisms for self- and other-knowledge of PAs, but only one (the mindreading faculty), the ISA theory is also simpler than any account of self-knowledge that assumes introspection. But simplicity considerations only come into play once there is indeed no evidence for introspective access to PAs. It might be suggested that we are able to know our own PAs introspectively via mental imagery such as inner speech or visual images.

Carruthers disagrees. He grants that imagistic states are non-interpretively accessible, for they are, just as sensory states, globally broadcast. But he contends that they don't themselves qualify as PAs (chapter 4). His main reason is that imagistic states don't play the right causal role to be PAs: while PAs such as judgments and decisions have the defining feature of settling matters and issuing directly into behavior, no kind of conscious mental imagery leads directly to action or settles what to do. For instance, suppose that upon reflecting on my habit to smoke, I decide to quit smoking and find myself say in inner speech "I'll stop smoking." In Carruthers' view, this inner speech utterance, while conscious and perhaps co-occurring with my decision to stop smoking, isn't itself the decision. For unless I also know what the words in the sentence mean, believe that this sentence applies to me, and have decided to implement what it says, uttering in inner speech "I'll stop smoking" won't lead to any behavior required for quitting to smoke. Carruthers claims that inner speech or any other instance of conscious mental imagery is only causally efficacious in virtue of the unconscious PAs that underlie the imagistic state. And since it lacks the causal efficacy of a PA, the imagistic state cannot itself constitute an attitude. Given that no instance of mental imagery constitutes a PA itself, introspective self-knowledge of mental imagery doesn't undermine the ISA theory, Carruthers holds.

In Carruthers' view, we only have interpretive access to our own PAs. But one might argue that this isn't always the case, for aren't we sometimes able to simply feel our own attitudes (e.g., the desire for a holiday, a cold beer, a loved one)? In response to this suggestion, Carruthers grants that the affective components of "affective attitudes," e.g., felt desires or likes/dislikes are globally broadcast, and can therefore be known introspectively. He maintains, however, that empirical evidence shows that we typically lack introspective knowledge of the particular property of an object that is responsible for our affective state. There is thus at best only limited non-interpretive self-knowledge of affective states.

Carruthers reports, for instance, a study by Johansson, Hall, Sikström, and Olsson (2005) in which subjects were shown pairs of photographs of female faces side by side, and were asked to quickly select the more attractive one. Shortly after their selection, the experimenter again presented what subjects took to be the selected photo, and asked them why they had chosen it. The photo that they were in fact shown, however, was the one they had just rejected. Most subjects didn't notice the switch, and started confabulating reasons for why they had chosen the face that they in fact initially rejected. If the subjects' affective responses to the faces on the photos had been tied to a particular property of the faces then, Carruthers holds, subjects should have been aware of the absence of that property after the switch. But this didn't happen. Johansson et al.'s (2005) findings indicate that subjects sometimes fail to recall the particular object properties that their affective PAs are tied to, and that they are liable to reconstruct them retrospectively.

This runs counter to most views of self-knowledge of PAs in the cognitive sciences. Typically, these theories invoke an "inner sense" faculty, i.e., a faculty that provides us with a direct quasi-perceptual channel of informational access to our own PAs, including their particular objects (Frith & Happé, 1999; Goldman, 2006; Nichols & Stich, 2003). These theories are challenged by the findings, for if we had such direct access to our own PAs, confabulations of attitudes and their objects should not occur.

In chapter 11, Carruthers mentions many more examples of attitude confabulation and self-interpretation that undermine inner sense accounts and more generally any theory assuming non-interpretive access to PAs. For instance, in an experiment by Brasil-Neto, Pascual-Leone, Valls-Solé, Gohen, and Hallett (1992), subjects failed to differentiate finger movements that they decided to perform from movements that were, unbeknownst to them, caused via stimulation of their motor cortex. Similarly, Wegner and Wheatley (1999) found that subjects could be made to self-ascribe intentions for actions that they did not perform merely by causing them to entertain action-relevant thoughts shortly before they witnessed the action effect. These findings suggest that against the subjects' own intuition, they don't have direct knowledge of their own PAs. Carruthers also mentions evidence that self-knowledge of PAs involves self-interpretation. For instance, subjects who have written an essay on a claim they disagree with shift their attitude when they unexpectedly receive a significant amount of money briefly before completing the attitude questionnaire after the task (Jordens & Van Overwalle, 2005). They appear to interpret their current positive mood (caused by receiving the money) as pertaining to the claim that they in fact just assessed negatively. Similarly, when subjects nod their heads while listening to an unpersuasive message, they are more likely to think they agree with the message (Briñol & Petty, 2003).

Carruthers holds that while the ISA theory predicts that confabulations and self-interpretations should occur, these and many other findings contradict any account that assumes introspective access to PAs. He concludes that, in the absence of any evidence for non-interpretive knowledge of PAs, we should assume that we lack non-interpretive knowledge of our own PAs. Furthermore, given that the ISA theory is well in line with the empirical evidence and other theories in the cognitive sciences

and also offers a simpler account of self-knowledge of PAs than any account postulating introspection, we should prefer the ISA theory to alternative proposals.⁴

3. The ISA Theory, Anarchic Hand Syndrome and Utilization Behavior

Even though I am sympathetic to Carruthers' theory of self-knowledge and his argument against introspection, there is a pathological case that challenges his ISA account. As noted, throughout his book, Carruthers cites various studies in which subjects unconsciously confabulate decisions and intentions to make sense of their behavior post hoc. These studies offer strong evidence against introspection of those PAs, and support his argument for the ISA theory. There are, however, also clinical cases in which subjects don't retrospectively construct intentions for their actions.

Consider, for instance, subjects with the so-called "anarchic hand syndrome" (AHS) (Della Sala, Marchetti, & Spinnler, 1991; Goldberg & Bloom, 1990). With one of their hands, AHS patients exhibit complex goal-directed movements that they can't voluntarily control. For instance, Della Sala (2005) mentions a case in which an AHS patient had dinner together with her parents

when, out of the blue and much to her dismay, her left hand took some leftover fish-bones and put them into her mouth . . . A little later, while she was begging it not to embarrass her any more, her mischievous hand grabbed the ice-cream that her brother was licking. (Della Sala, 2005, p. 606)

AHS patients are well aware that their involuntary movements are theirs. However, "they do not accept that the 'I' is the source of action," and do "not reconstruct a conscious intention to satisfy the logical role as the cause of the action that they accept as theirs" (Haggard, 2005, p. 294). If subjects acquire self-knowledge of PAs such as intentions via interpretation of their own behavior, however, one might wonder why AHS patients do not retrospectively confabulate intentions for their sudden odd movements and integrate them into their self-conception.

Interestingly, in another disorder, utilization behavior (UB), this does seem to happen (Bayne & Pacherie, 2007).⁵ UB patients exhibit similar, apparently involuntary stimulus-driven actions as subjects with AHS. For instance, they often can't resist suddenly reaching out and automatically using objects around them in an "object-appropriate" manner that is nonetheless inappropriate for the particular context (Archibald, Mateer, & Kerns, 2001; Lhermitte, 1983). For example, "a patient may pick up a toothbrush and begin to brush his teeth, in response to a toothbrush being placed in front of him, but in a context or setting in which brushing teeth would not normally be expected or done, such as in an appointment with a doctor" (Archibald et al., 2001, p. 119). In a different case, "a patient began to undress and get into a bed upon entering a room in which there happened to be a bed in plain sight" (Archibald et al., 2001, p. 119).

Even though the behavior of UB patients is, just as in AHS, compulsive, stimulus-driven, and contextually inappropriate, subjects don't view their own action as in any way strange. In fact, Lhermitte (1983) notes that UB patients who are asked about the

reasons for their odd behavior typically say that their actions had to be performed and were entirely normal. That is, unlike AHS patients, subjects with UB don't experience their behavior as unintended. A theory that holds that subjects come to know their PAs via interpreting their own overt behavior needs to account for the contrast between AHS and UB, for the behavioral symptoms in these disorders are similar while self-knowledge of intention differs significantly.

How could Carruthers' ISA theory account for these conditions? According to the ISA theory, in one's own case, the mindreading faculty doesn't only have access to one's own overt behavior, but also to mental imagery (e.g., visual images and inner speech), as well as sensory states, and emotional feelings. Carruthers might suggest that a difference in these inner states could be responsible for the difference in the self-ascription of intentions in UB and AHS patients. For instance, given the additional resources of data available to the mindreading faculty in one's own case, the fact that AHS patients don't self-ascribe the involuntary movements of their alien hand might be grounded in the presence or absence of a particular kind of mental imagery or feeling, indicating to the subject that these movements are not intended. Conversely, in the case of UB, this imagery might not be available to the mindreading faculty, resulting, unlike in AHS, in the subject's self-attribution of her odd behavior. So on this view, the mindreading faculty would be intact in AHS and UB patients, but it would receive different informational input pertaining to their inner states, leading to different interpretive output in both disorders.

On this proposal, UB and AHS patients no longer pose a problem for Carruthers' account. However, the proposal rests on the assumption that the mindreading faculty is well-functioning in AHS and UB patients. What reason do we have for believing that this assumption is true?

Consider AHS first. If AHS patients had a dysfunctional mindreading faculty then we would expect that they also have difficulties with making sense of the movements of their non-anarchic hand. But this is not the case. AHS patients don't have difficulties in recognizing the movements of their non-anarchic hand as intended (Pacherie, 2007). Furthermore, the fact that they don't self-attribute the movements of their anarchic hand indeed suggests positively that they have an *intact* mindreading faculty. Given that the faculty has access to all of the subject's behavior, it should detect a difference between the non-anarchic and the anarchic hand movements. Since the anarchic hand movements are very unlikely to cohere with the overall self-conception of AHS patients, the mindreading faculty of AHS patients should interpret these movements as caused by someone else (Pacherie, 2007). And this is, in fact, what we find. The symptoms of AHS are, thus, not only well compatible with a functioning mindreading faculty, but they seem to evidence it.

The same doesn't apply, however, when it comes to UB. There is neuroscientific and psychological evidence suggesting that patients with UB have a dysfunctional mindreading faculty. For instance, Baron-Cohen et al. (1994) note that people with damage in the orbitofrontal cortex (which is thought to play a significant role in mindreading, and is impaired in people with autism, who are known to have mindreading deficits) also have UB. Furthermore, while subjects with UB

“confabulate explanations for why they acted, their confabulations are unusual. . . . UB patients offer only impersonal rationalizations such as ‘I thought this was the thing to do’ or ‘I thought you wanted me to do that’” (Pacherie, 2007, p. 215). These subjects with UB don’t explain their odd actions in terms of their *own* PAs, even though they also don’t take them to be unintended. With the two assumptions that (1) they have a well-functioning mind-reading faculty and (2) subjects come to know their own PAs, including intentions, only via the mindreading faculty turned inwards, this behavior of UB subjects seems unusual indeed. Neuroscientific and psychological findings thus provide reasons to believe that subjects with UB, unlike AHS patients, have a deficient mindreading faculty.

The assumptions that the mindreading faculty is well-functioning in both AHS and UB, and that the difference in AHS and UB self-ascriptions of behavior and intentions is due to the mindreading faculty’s interpretation of different kinds of mental imagery or other inner states, are hence less plausible than assuming that the difference is due to an impaired mindreading faculty in UB subjects. Note that this account of AHS and UB would still be consistent with Carruthers’ ISA theory.

However, given the preceding considerations, a problem with his view arises with the following patient. Bayne and Pacherie (2007) and Marcel (2003) mention a subject who suffered from both AHS and UB simultaneously. The subject “exhibited Anarchic Behavior in his left hand and Utilization Behavior in his right hand, being troubled by the former but unconcerned about the latter” (Marcel, 2003, p. 78). If both disorders can occur in a subject at the same time, the just-mentioned explanation of the behavior of UB in terms of an impaired mindreading faculty and AHS in terms of an intact mindreading faculty no longer seems plausible. The subject that Marcel (2003) mentions contradicts this explanation, for if we assume that UB is characterized by a dysfunctional mindreading faculty, and that AHS indicates a well-functioning mindreading faculty, then a subject shouldn’t have both conditions at the same time.

The challenge that this case poses for the ISA theory can be summarized thus. Let’s assume that, as the ISA theory claims, the only way in which subjects come to know their intentions is via the mindreading faculty’s interpretation. If that were so, then either the mindreading faculty is well-functioning in AHS subjects or it isn’t. If the mindreading faculty is defective in AHS subjects then, as mentioned, we would expect that AHS subjects have difficulties in discriminating between intended and unintended hand movements.⁶ That is, they should have difficulties in noticing that their anarchic hand movements are unintended. But they don’t have such difficulties. Given this, we have reason to believe that the mindreading faculty is intact in AHS patients. However, if the faculty were intact then we would expect that Marcel’s patient, who has AHS and UB, should be able to detect a mismatch between his intended movements and his odd, unintended UB actions. This is not what we find. Thus, the assumption that the interpretation by the mindreading faculty is the only way in which a subject can come to know her own Pas, including intentions, leads to the problem that the faculty is both functional and dysfunctional in a subject (Marcel’s patient) at the same time. Since this is a contradiction, we should reject the

assumption that leads to it, namely that mindreading is the only way in which subjects can come to know their own intentions.

Given that the ISA theory claims that there is just one mechanism by which subjects come to know their own PAs, namely the mindreading faculty, Carruthers now needs to provide a different, ISA-theory-compatible account of UB and AHS than the one provided above, which considers both the possibilities that these disorders might involve an intact and a dysfunctional mindreading faculty. It is not obvious how this can be done. One might respond to the preceding line of argument that, contrary to my assumption, one can account for AHS and UB without appealing to any kind of meta-cognition. All that is required to explain these disorders and the odd self-attributions of intentions involved is an appeal to a first-order self-monitoring mechanism that might be intact or broken.

Indeed, while Carruthers doesn't discuss AHS and UB, he does consider other abnormal cases of agency experiences and self-attribution of intention. For instance, he discusses subjects with passivity forms of schizophrenia, who report thought insertion and an absence of control over their own actions. Maybe he could explain AHS and UB in the same way as he explains this condition. Carruthers argues that the passivity symptoms in these schizophrenic patients (e.g., feelings of alien control) are the result of a failure in self-monitoring that is first-order in nature; i.e., it is not due to a dysfunctional meta-cognitive monitoring of one's own mental states (2011, p. 296). More specifically, he holds that a failure in the comparator mechanism involved in the control of action (Frith, Blackmore, & Wolpert, 2000) leads to passivity forms of schizophrenia. In normal cases, the mechanism receives a forward model of the expected sensory feedback (efference copy) of an intended movement, and compares it with the incoming sensory feedback (reafference). Depending on a match or mismatch, it then adjusts motor plans to ensure a successful execution of the action (Jeannerod, 2006).

There is good evidence that this comparator mechanism is damaged in schizophrenics with passivity symptoms. For instance, while in normal subjects the mechanism accounts for the inability to tickle oneself (based on knowledge about a subject's own hand movements, it expects a certain sensory stimulation that is then "cancelled out"), schizophrenics with passivity symptoms are able to tickle themselves (Blackmore, Smith, Steel, Johnstone, & Frith, 2000). In contrast to normal subjects, schizophrenics experience their own movements as sensorily vivid, as if someone else were making their movements for them, and they will experience their own inner speech just as if another person was speaking (Carruthers, 2011, p. 296). A faulty comparator mechanism provides an explanation for why schizophrenics believe that their own action is somebody else's and that some of their thoughts have been inserted into their minds.⁷

However, unlike in these cases of schizophrenia, in AHS and UB subjects, the comparator mechanism appears largely intact. For instance, they don't experience their own movements as caused by someone else (compare: *alien* hand syndrome, a different disorder from AHS, in which subject *do* experience their own hand and movements as theirs; Marchetti & Della Salla, 1998), and also don't experience

thought insertion. There are thus important differences between AHS and UB, on the one hand, and schizophrenia, on the other, that make appealing to a dysfunctional comparator mechanism to explain AHS and UB less promising than in the case of passivity forms of schizophrenia.

There might be other ways of explaining AHS and UB by appeal to first-order self-monitoring mechanisms rather than a meta-cognitive faculty. My point here is not that such an explanation isn't possible, but only that these conditions, where they occur together, pose an interesting challenge to the ISA theory.

4. Motor Cortex Stimulation and Conscious Intentions

Let me now turn to a point of criticism concerning Carruthers' use of a particular experimental finding in his case against introspection. In chapter 11 and various previous publications (2009a, 2009b, 2010), Carruthers cites, as mentioned above, experimental findings by Brasil-Neto et al. (1992). Brasil-Neto et al. conducted an experiment involving focal magnetic stimulation of areas of the motor cortex. Subjects were instructed that when they heard a click, which was in fact the sound of the magnet being activated, they should lift one or another index finger. They could choose which finger to move. Areas of the motor cortex were then stimulated, either on the right or the left. In such cases, the subjects exhibited a strong tendency to raise the index finger of their contralateral hand. Unbeknownst to the subjects, their finger movement was due to the stimulation of their brain. But, Carruthers writes, "in each case they claimed to have been aware of *deciding* to lift that finger" (2011, p. 334). Carruthers argues that these findings contradict introspection-based views and lend support to his ISA theory. He holds that:

Since from the perspective of the subjects' mindreading systems the best explanation of the available data is that they chose to lift the index finger that subsequently moved, that is what they report. But they are unaware that they make these reports as a result of self-interpretation. (2011, p. 334)

It is not obvious that the subjects in Brasil-Neto et al.'s experiment did indeed engage in self-interpretation. An alternative explanation of the findings would be that the motor cortex stimulation not only caused the finger to move but, prior to that, also produced a desire to move the finger that the subject then correctly reported after the movement occurred. Carruthers rejects this possibility, writing that:

Everything we know about the organization of the brain suggests that motor cortex is not the place where decisions themselves are located. On the contrary, frontal cortex is generally thought to be the seat of such executive functions. Nor is there any plausible route via which stimulating motor cortex could cause a decision to be made. . . . So the cranial stimulation is unlikely to have caused a decision to lift the contralateral finger to come into existence, which was then introspected. (2011, pp. 333–334)⁸

There are two points to note here. First, Carruthers holds that the test subjects made “reports” and “claimed to have been aware of *deciding* to lift” the finger that they did lift upon cranial stimulation (2011, p. 334). However, in fact, Brasil-Neto et al. nowhere in their paper mention what the subjects claimed with regard to their finger movement. They only write that 80% of subjects “were not aware of any effect of [the magnetic stimulation of their motor cortex] on their response pattern” (Brasil-Neto et al., 1992, p. 965). Subjects’ not noticing that their finger movement was in fact caused by motor cortex stimulation is one thing; their claiming that they decided to move their finger is another. For all the paper says, only the first actually happened in the experiment.

Second, Carruthers claims that decisions are unlikely to be caused by motor cortex stimulation. However, transcranial stimulation in one area of the cortex can in fact have very rapid (1 synapse, or 10 ms) effects in other areas. Given that cortex activation can spread rapidly, the assumption that only the directly stimulated area is being activated remains questionable.⁹ Furthermore, even if we agree with Carruthers that *decisions* are not caused by motor cortex stimulation, it might still be that motor cortex stimulation gives rise to a more basic want or urge to lift the index finger. Indeed, as it turns out, a study by Fried et al. (1991) provides evidence in support of such proposal. In Fried et al.’s study, patients received direct stimulation of the supplementary motor area (SMA) via two strips of electrodes. During low intensity stimulation of the SMA, the subjects reported a want or urge to move a specific body part. When the stimulation intensity was increased, actual movement of the same body part occurred. Fried et al.’s experiment suggests that a conscious state “akin to intention” is “a direct product of the neural processes in the SMA that generate movement” (Haggard, 2005, p. 293). Given that SMA stimulation can generate a wanting to act in a certain way, there are good grounds for assuming that the same happened in Brasil-Neto et al.’s experiment, in which case the subjects’ reports that they wanted to move the finger (assuming they did indeed make such reports) would turn out to be veridical. Brasil-Neto et al.’s findings would no longer bolster Carruthers’ argument against introspection.

Indeed, Fried et al.’s findings offer hints towards a different account of self-knowledge of intentions—at least of intentions to actions—than Carruthers’ ISA theory. A number of studies (e.g., Haggard & Clark, 2003; Lau, Rogers, Haggard, & Passingham, 2004; Sirigu et al., 2004) suggest that the frontal and parietal lobes together “form a circuit which elaborates and monitors motor plans in advance of action, producing a conscious experience of intention as part of this simulation” (Haggard, 2005, p. 292). On this account, conscious intention is an experience that might involve retrospective interpretation, but it is also generated as a result of pre-movement neural activity in the frontal and parietal motor areas. If intentions to act are sometimes directly experienced then the intuition of introspective access to one’s own PAs, at least to one’s own intentions to act, may after all not only be a useful illusion of the mind to simplify its processing, as Carruthers proposes (chapter 1), but rather be grounded in the way we experience ourselves as agents.

Acknowledgments

Many thanks for comments and helpful discussions to Nick Shea, Matteo Mameli, and Charlie Leach.

Notes

- [1] For Carruthers, these states also include sensorily embedded judgments (i.e., judgments such as *I see a red tomato* that are directly grounded in sense perception, e.g., vision), and felt desires.
- [2] Note that only the nature of self-knowledge of *conscious* PAs (e.g., active, occurrent judgments such as *It looks like it isn't going to rain*) is at issue. By 'conscious', Carruthers means *access* conscious (see Block, 1995, for the distinction). It is uncontroversial that standing PAs or Freudian beliefs and desires are only known interpretively.
- [3] See Gazzaniga (1995) for details on the condition.
- [4] In the last chapter of his book, Carruthers draws some sweeping conclusions from his case for the ISA theory. He holds that the ISA theory, in conjunction with what he claims to be the two most plausible theories of attitude consciousness, entails that there are no conscious PAs, and that this, in turn, undermines the folk notions of the self, agency, and moral responsibility. For reasons mentioned elsewhere (Peters, forthcoming, unpublished manuscript), I find his arguments for these claims unconvincing. I shall ignore them here.
- [5] But see Marcel (2003, p. 76) for a discussion of differences between the two disorders.
- [6] Indeed, autistic subjects, who are known for their impairment in mindreading, also have difficulties identifying which action was caused by themselves and is in line with their own intentions (Frith & Happé, 1999; Lang & Perner, 2002; Phillips, Baron-Cohen, & Rutter, 1998).
- [7] However, note that on this account one would expect that schizophrenics with passivity symptoms also exhibit impaired ordinary motor learning. They don't do so. This provides a reason to be skeptical of the faulty-comparator explanation of schizophrenia.
- [8] Presumably, by 'frontal', Carruthers means 'prefrontal' in this passage. For the prefrontal cortex, but not the frontal cortex in general, excludes the motor strip.
- [9] Thanks to Nick Shea here.

References

- Archibald, S.J., Mateer, C.A., & Kerns, K.A. (2001). Utilization behavior: Clinical manifestations and neurological mechanisms. *Neuropsychology Review*, *11*, 117–130.
- Baars, B. (1988). *A cognitive theory of consciousness*. New York, NY: Cambridge University Press.
- Baddeley, A., & Logie, R. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 28–61). Cambridge: Cambridge University Press.
- Baron-Cohen, S., Ring, H., Moriarty, J., Schmitz, B., Costa, D., & Ell, P. (1994). Recognition of mental state terms: Clinical findings in children with autism and a functional neuroimaging study of normal adults. *British Journal of Psychiatry*, *165*, 640–649.
- Bayne, T., & Pacherie, E. (2007). Narrators and comparators: The architecture of agentive self-awareness. *Synthese*, *159*, 475–491.
- Blakemore, S.-J., Smith, J., Steel, R., Johnstone, E.C., & Frith, C.D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychological Medicine*, *30*, 1131–1139.

- Block, N. (1995). A confusion about the function of consciousness. *Behavioral and Brain Sciences*, 18, 227–247.
- Brasil-Neto, J., Pascual-Leone, A., Valls-Solé, J., Gohen, L.G., & Hallett, M. (1992). Focal transcranial magnetic stimulation and response bias in a forced choice task. *Journal of Neurology, Neurosurgery, and Psychiatry*, 55, 964–966.
- Briñol, P., & Petty, R. (2003). Overt head movements and persuasion: A self-validation analysis. *Journal of Personality and Social Psychology*, 84, 1123–1139.
- Byrne, R., & Whiten, A. (Eds.). (1997). *Machiavellian intelligence II*. Cambridge: Cambridge University Press.
- Carruthers, P. (2009a). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, 32, 121–138.
- Carruthers, P. (2009b). Simulation and the first-person. *Philosophical Studies*, 144, 467–475.
- Carruthers, P. (2010). Introspection: Divided and partly eliminated. *Philosophy and Phenomenological Research*, 80, 76–111.
- Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge*. Oxford: Oxford University Press.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79, 1–37.
- Della Sala, S. (2005). The anarchic hand. *The Psychologist*, 18, 606–609.
- Della Sala, S., Marchetti, C., & Spinnler, H. (1991). Right-Sided anarchic (alien) hand: A longitudinal study. *Neuropsychologia*, 29, 1113–1127.
- Dennett, D. (1992). The self as a center of narrative gravity. In F. Kessel, P. Cole, & D. Johnson (Eds.), *Self and consciousness: Multiple perspectives* (pp. 103–115). Hillsdale, NJ: Erlbaum.
- Farkas, K. (2008). *The subject's point of view*. Oxford: Oxford University Press.
- Fried, I., Katz, A., McCarthy, G., Sass, K.J., Williamson, P., Spencer, S.S., & Spencer, D.D. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. *Journal of Neuroscience*, 11, 3656–3666.
- Frith, C., Blackmore, S., & Wolpert, D. (2000). Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews*, 31, 357–363.
- Frith, U., & Happé, F. (1999). Theory of mind and self-consciousness: What is it like to be autistic? *Mind and Language*, 14, 1–22.
- Gathercole, S., & Baddeley, A. (1993). *Working memory and language*. Hillsdale, NJ: Erlbaum.
- Gazzaniga, M. (1995). Consciousness and the cerebral hemispheres. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1391–1399). Cambridge, MA: MIT Press.
- Goldberg, G., & Bloom, K. (1990). The alien hand sign. Localization, lateralization and recovery. *American Journal of Physical Medicine & Rehabilitation*, 69, 228–238.
- Goldman, A. (2006). *Simulating minds*. Oxford: Oxford University Press.
- Haggard, P. (2005). Conscious intention and motor cognition. *Trends in Cognitive Sciences*, 9, 290–295.
- Haggard, P., & Clark, S. (2003). Intentional action: Conscious experience and neural prediction. *Consciousness and Cognition*, 12, 695–707.
- Hirstein, W. (2005). *Brain fiction: Self-Deception and the riddle of confabulation*. Cambridge, MA: MIT Press.
- Jeannerod, M. (2006). *Motor cognition*. Oxford: Oxford University Press.
- Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, 310, 116–119.
- Jordens, K., & Van Overwalle, F. (2005). Cognitive dissonance and affect. *Psychologica Belgica*, 45, 157–184.
- Lang, B., & Perner, J. (2002). Understanding of intention and false belief and the development of self-control. *British Journal of Developmental Psychology*, 20, 67–76.
- Lau, H.C., Rogers, R., Haggard, P., & Passingham, R. (2004). Attention to intention. *Science*, 303, 1208–1210.

- Lhermitte, F. (1983). Utilization behavior and its relation to lesions of the frontal lobes. *Brain*, 106, 237–255.
- Marcel, A.J. (2003). The sense of agency: Awareness and ownership of actions and intentions. In J. Roessler & N. Eilan (Eds.), *Agency and self-awareness* (pp. 48–93). Oxford: Oxford University Press.
- Marchetti, C., & Della Salla, S. (1998). Disentangling the alien and anarchic hand. *Cognitive Neuropsychiatry*, 3, 191–208.
- Nichols, S., & Stich, S. (2003). *Mindreading*. Oxford: Oxford University Press.
- Pacherie, E. (2007). The anarchic hand syndrome and utilization behavior: A window onto agentic self-awareness. *Functional Neurology*, 22, 211–217.
- Peters, U. (forthcoming). Conscious propositional attitudes and moral responsibility. *Journal of Moral Philosophy*.
- Peters, U. (unpublished manuscript). Self-Knowledge, and conscious propositional attitudes.
- Phillips, W., Baron-Cohen, S., & Rutter, M. (1998). Understanding intention in normal development and in autism. *British Journal of Developmental Psychology*, 16, 337–348.
- Ryle, G. (1949). *The concept of mind*. London: Hutchinson.
- Searle, J. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge: Cambridge University Press.
- Shoemaker, S. (1996). *The first-person perspective and other essays*. Cambridge: Cambridge University Press.
- Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., & Haggard, P. (2004). Altered awareness of voluntary action after damage to the parietal cortex. *Nature Neuroscience*, 7, 80–84.
- Stephens, G., & Graham, G. (2000). *When self-consciousness breaks: Alien voices and inserted thoughts*. Cambridge, MA: MIT Press.
- Wegner, D., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54, 480–491.