*Philosophical Papers*
Vol. 38, No. 2 (July 2009): 133-156

# Action-Awareness and the Active Mind

*Peter Carruthers*

**Abstract:** In a pair of recent papers and his new book, Christopher Peacocke (2007, 2008a, 2008b) takes up and defends the claim that our awareness of our own actions is immediate and not perceptually based, and extends it into the domain of *mental* action.[1] He aims to provide an account of action-awareness that will generalize to explain how we have immediate awareness of our own judgments, decisions, imaginings, and so forth. These claims form an important component in a much larger philosophical edifice, with many implications for the philosophy of mind and for epistemology. The present paper advances multiple criticisms of Peacocke's account. In particular, it shows that he has given insufficient grounds for thinking that we have non-perceptual awareness of our own actions; and it shows that the account of action-awareness that he provides does not, in any case, generalize to mental actions of the sort that he intends.

## 1. Introduction

Christopher Peacocke's proximal goal is to explain how we have immediate knowledge of such mental events as judging, deciding, reasoning, imagining, and so forth. (Ultimately, he wants to establish a particular account of the 'possession conditions' of mental concepts, one that implicates immediate first-person reference to instances of those concepts.) He therefore needs to demonstrate that we can possess such knowledge in a way that is independent of any of the sensory or quasi-sensory accompaniments of the events in question (such as 'inner speech' and visual imagery). In pursuit of these goals he aims to establish three claims. The first is that propositional attitude events of the sort listed above are genuinely a species of *action*, belonging to the same general category as our physical actions. The second is what he calls his

1 There is extensive overlap between the two articles and Chapter 7 of Peacocke's book (2008b), whereas the book contains some relevant discussion (in Chapter 6) not included in the articles. Unless otherwise indicated, therefore, all references to Peacocke should be understood as citing his 2008b.

Routledge
Taylor & Francis Group

ISSN 0556-8641 print/ISSN 1996-8523 online

DOI: 10.1080/05568640903146443
http://www.informaworld.com

UNISA university of south africa PRESS

'Principal Hypothesis', which is that our *awareness* of our own mental actions is a species of action-awareness, being of the same kind as our awareness of our physical actions. And the third is that our awareness of our own physical actions is immediate and independent of any form of sensory awareness.

It is important to stress that the overall form of the argument is intended to be an inference to the best explanation (as Peacocke himself emphasizes). The Principal Hypothesis is put forward *because* it is thought to do the best job of *explaining* how we have immediate awareness of judging, deciding, and so forth. And likewise the Principal Hypothesis itself lends support to the other main claims involved—in particular, that we have immediate awareness of both mental and physical actions. (As is familiar, the confirmation-relation is holistic.) This means that the Principal Hypothesis cannot be interpreted merely phenomenologically, at the personal level of description. For suppose that the hypothesis that our awareness of our physical and mental actions is of the same kind amounted to no more than the claim that in each case we have knowledge that is not grounded in sensory awareness. Then the *explanans* would amount to nothing more than a conjunction of the *explanandum* together with a similar claim made in respect of physical actions. The intended inference to the best explanation would therefore fail. (You cannot use inference to the best explanation to argue for the truth of $P$ & $Q$ on the grounds that its truth explains the truth of $P$. Nor does the supposition that $P$ & $Q$ lend any additional support to the hypothesis that $P$.)

The Principal Hypothesis must therefore be interpreted at the subpersonal level, in terms of shared mechanisms. Since the hypothesis that there is a single mechanism of action-awareness would genuinely *explain* why our knowledge of both physical and mental actions is immediate, an inference to the truth of that hypothesis is at least a *candidate* inference to the best explanation. (The extent to which any such inference succeeds will depend on the strength of the alternatives, of course.) Moreover the plausibility of that hypothesis will, in turn, lend

support to the claim that we have immediate awareness of our mental actions. And indeed, just as one would predict, Peacocke does make claims about the subpersonal mechanisms of action-awareness (while also putting those claims to work in explaining the co-occurrence of symptoms in schizophrenia, as we shall see).

Given Peacocke's overall goals, however, it is quite crucial that he should be able to establish the third of the three claims listed above, that our awareness of our own physical actions is independent of any sensory or quasi-sensory (e.g., imagistic) data. For otherwise, showing that our awareness of our mental actions is of the same general type as our awareness of our physical actions wouldn't help in establishing that we have awareness of our mental actions that is independent of sensory or quasi-sensory events. This is where we begin, in Section 2. I shall argue that Peacocke has failed to make even this basic aspect of his case. Later sections will then proceed to discuss the other main features of his argument.

**2. Action awareness**

One of Peacocke's arguments is this: we can know of our own movements even when the relevant portion of our body is anaesthetized and transmitting no kinesthetic or other somatosensory information. At the dentist, for example, with one's entire jaw completely 'frozen', one can still know that one has just opened one's mouth. And one can know this immediately, without needing to see oneself in a mirror or hear the 'thank you' from the dentist for complying with his request. This example is unconvincing by itself, however, because opening one's mouth (especially opening it wide, as one does at the dentist) is associated with tightening of the muscles in one's neck as well as changes in the elasticity in the surrounding skin, in regions that have *not* been frozen by the local anesthetic. So it remains possible that one's knowledge of the action of opening one's mouth is grounded in sensory data of these sorts.

Peacocke also appeals to cases of people whose reafferent nerves in one or more limbs have been destroyed. These patients are no longer

capable of receiving somatosensory information from the relevant portions of their bodies, while retaining the capacity for movement. One such patient has lost all forms of feeling in one arm, but she can still tell, immediately and without having to look, that she is raising her arm or flexing her wrist—or so Peacocke tells us. Unfortunately, Peacocke provides no references for these claims, which appear to be contradicted elsewhere in the literature. Specifically, Jeannerod (2006) claims that such patients remain *un*conscious of their movements when deprived of visual access to them. We need to examine his argument.

Jeannerod describes an experiment in which the patient is instructed to draw a straight line towards a target. Both the target and the line drawn by the pen are displayed in a mirror placed over the subject's hand, obscuring the latter from view. The mirror reflects a video screen whose image is driven by the touch-sensitive pad on which the patient draws. But the relationship between the direction in which the patient actually needs to move the pen to secure the intended screen display can be manipulated by the experimenter, in such a way that to achieve a line that moves vertically towards the target the subject has to make a drawing-movement that is displaced by 10° from the vertical, say. Normal subjects make smooth and accurate corrections for small displacements of up to about 10°, without any awareness that they are doing so. Beyond that point they become aware that they are having to compensate, and their movements become much less smooth and accurate in consequence. But the patient with reafferent nerve damage failed to become aware of the changes in her movements even for displacements much greater than 10°, although she did report that the task became harder and more effortful, requiring greater concentration. In consequence, Jeannerod claims that she lacks awareness of her own actions when using the damaged limb in the absence of visual input.

Jeannerod's conclusion is too hastily drawn, however. For the case has some analogies with instances of change-blindness. Someone who views two similar pictures that alternate with one another can remain unaware that they differ in some significant respect. (Both may be pictures of the

same house, for example, but in one picture the windows are shuttered whereas in the other they are not.) But this does not entail that the subject is unaware of the respect in which the pictures differ. Someone can be conscious of the shutters in one picture and conscious of bare windows in the other, without being aware *that* the windows are shuttered in one and bare in the other. (Note that the latter is a comparative judgment, requiring the subject to make a comparison between two visual representations held in working memory, whereas the former two judgments are not.) Likewise, one might think, the patient with reafferent nerve damage might be aware of moving her arm forwards without being aware that she is moving that arm in a somewhat different direction from a few moments previously. (Here, too, the latter is a comparative judgment whereas the former is not.) And Jeannerod provides us with no data suggesting that the patient is unaware of moving her arm at all, in the absence of visual input.

Although the data are not really conclusive either way, I propose to grant Peacocke the following claim for purposes of argument: someone can know what they are doing in the absence of any perceptual input deriving from the action itself. This is not yet sufficient to establish the claim that our knowledge of our own actions is immediate. For there remains the possibility that our knowledge in the relevant instances is grounded, not in *perceptual* data, but in *quasi-perceptual* data, specifically, in *motor imagery*.[2]

The topic of motor imagery has been heavily researched in recent decades, and much is now known. (Jeannerod, 2006, provides an accessible review.) It is known that motor imagery is distinct from visual and other forms of imagery, for example, and that motor images play an important role in acquiring new skills, as well as in maintaining old ones in the absence of overt practice (e.g., because one is injured). As with other forms of imagery, motor images are centrally-caused representations that simulate the relevant sorts of perceptual state.

2 This point is also made against Peacocke by Prinz (2007).

Specifically, a motor image is a representation of the somatosensory perceptions that one *would* undergo if a given motoric action were executed, caused by the activation of the relevant motor schema. So even someone whose reafferent nervous pathways are destroyed would still be capable of experiencing motor images, generated by the actions that she performs. And this would provide a basis on which she can know what she is doing.

There are well-developed and widely-accepted theories of the cognitive systems that create motor images, and of their function (Wolpert and Kawato, 1998; Wolpert and Ghahramani, 2000; Grush, 2004). When a motor schema is activated it normally sends afferent signals to the relevant muscle groups. (In the case of purely imagistic action, these afferent signals are largely suppressed.) But at the same time an efference copy of those signals is created and sent to one or more emulator systems, which contain a model of the kinematics of the body. There they are transformed into what is now generally called a 'forward model' of the action. This is a representation of the somatosensory input that should be expected if the action were completed as intended. The forward model is received by a comparator system that also receives somatosensory information as input, facilitating fast online correction of the action as it proceeds. For the most part this process is a relatively encapsulated one, and the motor images that get created by the emulator system from the efference copy of the motor instructions remain unconscious. But when attended to (as they would be whenever subjects are asked to report what they are doing) they can be 'globally broadcast' to the central systems responsible for belief-formation, self-attribution, and verbal report, thereby becoming conscious.[3]

Motor images have one further function that is worth mentioning here, since it will become important later. This is that they are often

3 For evidence of the existence of a global broadcasting architecture for perceptual and quasi-perceptual states, and of its close association with consciousness, see Baars, 2002; Dehaene and Naccache, 2001; Dehaene *et al*., 2003; Baars *et al*., 2003.

taken as input by other sorts of emulator system and used to create imagery within other sense modalities. When mentally rehearsing an action that I am about to perform, for example, I can use that rehearsal to generate a *visual* image of myself undertaking that same action. This, too, can be globally broadcast and taken as input by a variety of conceptual systems, leading one to make predictions about the likely physical or social effects of the action, for example. Indeed, there is evidence of the crucial role of motor cortex in manipulations and transformations of images of all types, including visual ones (Kosslyn, 1994; Ganis *et al*., 2000; Richter *et al*., 2000; Kosslyn *et al*., 2001; Lamm *et al*., 2001).

I have suggested, against Peacocke, that it may be our motor images of our own actions that enables us to have knowledge of those actions in the absence of any perceptual input. But now the strange thing is that Peacocke, too, appeals to the very same neuroscientific data on efference copies and forward modeling when attempting to ground his own position. (He uses the language of 'corollary discharge', rather than 'efference copy', but the idea is the same.) He thinks that it is via the efference copy of the motor instructions for an action that we have immediate and non-inferential knowledge of the action. He writes:

> When there is no corollary discharge, there is no action-awareness of the movement in question as one of your own actions, and this applies quite generally, both in bodily and in mental cases. (2008b, p.278.)

But how could this work, except via our awareness of a motor image of the action? It might be suggested that the efference copy is received as input by the self-attribution system *directly* (in addition to being sent to the emulator system where it is used to create a motor image, which generally remains unconscious). This might provide the self-attribution system with the information that it needs, in order to generate knowledge of the action being performed, in the absence even of a motor image of the action, let alone any sensory input. But this idea is surely a non-starter. For the efference copy is a copy of a set of motor instructions sent to the muscles that control movement. It will consist of

a long and complex description of the form, 'Tighten such-and-such muscle to such-and-such a degree; relax so-and-so muscle completely', and so on. This is not the right *sort* of representation, or in the right kind of format, to issue in knowledge of the action that would result. It would *have* to be taken as input by an emulator system containing a model of the kinematics of the body in any case, where it could be transformed into a representation of the expected bodily movement. But we already know that such an emulator exists, and that it issues in motor imagery. Postulating yet another instance of emulator system specifically to deliver immediate knowledge of the action being undertaken would be otiose.

Alternatively, it might be suggested that the unconscious motor-images that are normally used for fine-grained control of action are *also* routinely made available to the systems that issue in beliefs and verbal reports about our own actions, only without becoming conscious. This would enable Peacocke to preserve his claim that our knowledge of our own actions is immediate, since all that would figure in consciousness would be the sense that a certain sort of action is being performed, independently of any other sensory or quasi-sensory awareness. This is certainly conceivable. But to make good on the idea, one would have to claim that there is a direct channel—independent of global broadcast—between the systems that create forward models of action and the self-attribution system. For we have every reason to think that the contents of global broadcasts are always conscious. We have no independent grounds for believing in such a channel of information, however, whereas there is accumulating evidence that global broadcast is generally a necessary condition for a content to become reportable.[4]

I claim, then, that Peacocke has failed to make out his case that knowledge of our own actions is immediate and not perceptually or quasi-perceptually grounded. Indeed, there is good reason to think that such knowledge is grounded in our awareness of our own motor

4 See the references cited in footnote 3.

imagery, at least in cases where somatosensory perceptual cues are not available. But let me set this criticism to one side for the moment, and move on to consider the remainder of Peacocke's argument.

**3. Mental actions**

Peacocke claims that mental events of judging, deciding, reasoning, and so on are genuinely a species of *action*. And as such, he thinks that they give rise to efference copies that provide us with immediate and non-sensory access to them. I shall consider the former thesis in the present section, deferring discussion of the latter to Section 4. The criterion that Peacocke uses, in order for an event to count as an action, is that it should constitutively involve a *trying*. An action is something that one can try, but fail, to do. This seems very natural in respect of most of the items in Peacocke's list. One can speak of trying to decide something, or of trying to imagine something. Likewise one can speak of trying to judge whether one line is longer than another, or whether polar bears are endangered.

In contrast, however, it would be very odd to speak of trying to wonder whether there will be a storm this afternoon, or of trying to suppose that there will be. Indeed, neither wonderings nor supposings figure explicitly in Peacocke's (open-ended) list of mental actions—and it seems for good reason. But he is thereby confronted with a dilemma. For pre-theoretically, wonderings and supposings seem to be no less central to the category of mental action than are imaginings and decidings. Hence Peacocke must either explain why they are not really mental actions after all; or he must explain why (unlike other mental actions) one cannot seem to try but fail at them.

It is important to note, moreover, that the acceptability of the locution, 'I am trying to Φ' does not entail that the trying is *constitutive* of the Φ-ing. Consider, 'I am trying to become famous' or, 'I am trying to die.' These should really be glossed as saying that I am trying to do things that are likely to result in me becoming famous, or in my death. (Of course I can try to kill myself, and killing myself would be an action;

but dying is not.) Closer to home, one can try (but fail) to believe something, and one can try (but fail) to forget something. But it would be *highly* implausible to claim that believing and forgetting are themselves mental *actions*. Rather, they are states or events that we can try to bring about in ourselves by action. Likewise, then, with, 'I am trying to judge whether the lines are of equal length.' It may be that the judging is not itself an action, but is rather an event that often results from actions. The trying might consist in looking more closely, asking myself the question, 'Are the lines really equal?', measuring them with a ruler, and so forth, none of which is constitutive of the judgment in question.

So while I am happy to accept the criterion that an action is an event that constitutively involves a trying, I want to emphasize that the fact that we describe ourselves as trying to decide, or trying to imagine, does not yet settle the active status of the attempted events. There is a further theoretical issue arising, as to whether the tryings in question are genuinely *constitutive* of the events described, or are merely instrumental in bringing them about. In my view this issue can only be resolved by considerations of cognitive architecture, taken together with evidence from cognitive science. It cannot be resolved by introspection, since no one thinks that introspection is sensitive to the difference between causation and constitution. Nor can it be resolved by reflection on common-sense concepts or common-sense ways of thinking and speaking, since these can only tell us about our pre-theoretical commitments, not about what is really the case.

**4. Awareness of mental actions**

Supposing, for the moment, that propositional attitude events of judging, deciding, and so forth are genuinely actions, we can ask how we have first-person knowledge of them. Peacocke's answer is: in the same way that we have knowledge of our own physical actions: immediately and without needing to rely upon sensory data, via the efference copies that they generate. One consideration offered in support of this proposal

is that it provides us with a unified account of our knowledge of both our mental and our physical actions. A second consideration is that the account provides the best explanation of the phenomenology of mental-action awareness. I shall consider these arguments in turn, before turning to evaluate the thesis itself.[5]

The unification-argument is perfectly acceptable so far as it goes. But it does not take us very far. In general it is reckoned to be a good thing if we can demonstrate that two apparently distinct sets of phenomena are really one and the same, or if we can show that they have a common underlying basis. Hence, other things being equal, the fact that a given proposal would unify two sets of phenomena is a reason for accepting that proposal. But two qualifications are in order. First, the proposal in question must actually be *successful* in accounting for both sets of phenomena. Just such a claim will be critically examined in respect of Peacocke's unifying thesis shortly. And second, theoretical simplification is only one good-making feature of a theory amongst others. Hence even a theory that would successfully unify two sets of phenomena, if it were true, might be defeated on other grounds.

As for the second argument offered in support of his position, Peacocke is correct that his account can explain the phenomenology of mental action-awareness. If we grant that people have immediate knowledge of their physical actions by virtue of the efference copies that those actions generate, and if we grant that our knowledge of our mental actions has exactly the same sort of basis, then of course we can explain why subjects should have immediate knowledge of their own judgments, decisions, and so forth. But whether this account provides the *best* explanation of the phenomena is moot. One competitor account is that we have access to our own propositional attitudes via the operations of one or more *sui generis* introspective mechanisms, which are distinct from the mechanisms that allow us to access our own physical actions

5 Peacocke also provides a third argument in support of his Principal Hypothesis, relating to the explanation of some characteristic symptoms of schizophrenia. This will be considered in Section 5.

(Nichols and Stich, 2003; Goldman, 2006). Against this, Peacocke can perhaps claim that his own theory appeals to mechanisms that we have independent reason to believe in (namely efference-copying), whereas this competitor view does not.

A second contending account is that we know of our own propositional attitude events, as well as our own actions, via a swift and unconscious process of self-interpretation of the sensory and quasi-sensory data (Gazzaniga, 1995, 2000; Wegner, 2002; Wilson, 2002; Carruthers, 2009a, 2010). For on this account, too, it would of course *seem* to subjects that their knowledge of their own attitudes is immediate, since they have no awareness of the inferential processes that generate it. Moreover, since the mode of operation of the self-interpretation system is unconscious, it will issue in a bare phenomenology of a *seeming* that a given action (whether physical or mental) is being performed (which is not yet a belief that it is). And this *seeming* will be capable of grounding indexical thoughts about the events in question, of the form, '*This* action' or, '*This* judgment', just as Peacocke stresses in support of his own account. In addition, the present account, too, appeals to mechanisms that we have independent reason to believe in. For these are the very same mechanisms that we use when attributing mental states and intentional actions to other people (often called 'mindreading').

I shall return to consider these contrasts between Peacocke's account and its two main competitors in due course. But I want now to turn to direct evaluation of the strengths and weaknesses of that account itself. Recall that mental actions, like physical actions, are supposed to be known of via the efference copies of themselves that they create. But how can this be so? The efference copy of a physical action is a copy of a set of *motor* instructions. But when I make a judgment or take a decision, what is the efference copy supposed to be a copy *of*? There is no motor schema for judging that there will be a storm this afternoon, nor are any commands sent to one's muscles when one does. And likewise no physical movements of any sort are constitutive of deciding not to go to the pool. Even if judgments and decisions are a species of action, they

just are not the right *kinds* of action to generate an efference copy.

It might be replied that actions can be specified with varying degrees of abstractness. Consider, for example, the relationships between *paying the check*, *signing one's name*, and *making such-and-such movements with one's hand*. Perhaps judgments and decisions are like the most abstract level of description of a physical action, but differing in that they do not have to be implemented in any particular physical movement. Even if true, this will not really help, however. For it remains the case that only physical movements generate an efference copy and subsequent forward model of the executed action. The cognitive scientists who have developed such accounts distinguish between the goal, or intention, with which an action is executed (e.g., *paying the check*) and the *motor intention* that is selected to achieve the goal. Only the latter is supposed to generate an efference copy. The rationale is that forward models have the function of monitoring the performance of an action as it unfolds, swiftly correcting if the incoming perceptions of the movement do not match those encoded in the forward model.

It should not be denied, of course, that there exist mental actions with a motoric component. For as we noted earlier, there is evidence that motor cortex has a large role to play in manipulating and transforming visual images. And we also know that during episodes of inner speech not only the language comprehension areas of the brain but also language *production* areas and associated regions of motor cortex are active (Paulescu *et al*., 1993; Shergill *et al*., 2002). Hence efference copies might well play a role in our awareness of mental actions of imagined-seeing or imagined-speaking. (Indeed, they almost certainly do. For it will be the efference copy of a motor schema—in circumstances where the overt expression of the latter is suppressed—that is used to generate a forward model of the action, which in turn issues in conscious imagery.) But we do not need to appeal to any special capacity for introspective action-awareness to explain our access to such events. On the contrary, the resulting images will be available to whatever systems subserve the attribution of mental states to other people, since these will

require access to the outputs of perceptual systems to perform their functions. (Hence they will also have access to imagistic events, given that these share mechanisms with perception, and given that images are 'globally broadcast'.) And any access that these images might give us to underlying events of judging, deciding, and so forth would plainly *not* be independent of sensory data. Quite the contrary.

In response to these difficulties, one option for Peacocke would be to say that our awareness of our mental actions might not use the *exact same* mechanism employed for awareness of physical actions, but rather something analogous. Hence it is not via an efference copy that one knows that one is judging that there will be a storm, but rather via something that plays the same sort of role as an efference copy. Perhaps this is so. But this cannot be Peacocke's view. For it would leave him with the claim that there are two distinct but similar mechanisms for providing self-knowledge of actions, one for knowledge of our own physical actions and one for knowledge of our mental ones. And that would undermine the explanatory argument in support of Peacocke's Principal Hypothesis, together with its claimed unification of the two sorts of self-awareness.

A second option would be for Peacocke to drop any reliance upon efference copies in formulating his position, and to restrict himself to the claim that it is the very same introspective mechanism (whatever that might be) that gives us non-inferential knowledge of both our own physical actions and our own mental actions (of judging, deciding, and so forth). Although this would avoid the difficulties outlined above, it is still problematic. For of course any such mechanism would need to be physically realized in the brain. Yet we know that very different areas of the cortex are constitutively involved in action-control and in decision-making, in particular; and we know that the sorts of mental representations involved in such events will differ greatly from one another. Roughly, the processes that generate and control physical actions take place in pre-motor and motor cortexes, and involve nonconceptual representations of the muscle and other changes that are

needed to issue in a given action; whereas the key aspects of decision-making take place in the frontal lobes, and are fully conceptual in character. It is quite mysterious how there could be any one mechanism that monitors and provides awareness of events of such disparate sorts, taking place in two distinct brain regions. Indeed, it is considerations of just this kind that lead Nichols and Stich (2003) to postulate multiple introspective self-monitoring mechanisms for perception, for action, and for each of the basic propositional attitude types.

A final option open to Peacocke would be to abandon all appeals to underlying mechanisms in formulating his Principal Hypothesis. He could restrict himself to the claim that our awareness of our mental actions is like our awareness of our physical actions in that both are phenomenologically immediate and are not grounded in prior forms of sensory awareness. This might be enough to subserve some of his purposes. (In particular, his account of the possession conditions for mental concepts might still be defensible.) But notice that the resulting claims would be quite independent of one another. Hence even success in establishing that our awareness of our own physical actions is immediate would lend no support for the parallel view in respect of mental action (or vice versa). And the claim that events of judging, deciding, and so forth are genuinely a species of action could be dropped altogether, since it would no longer be doing any explanatory work. Despite these drawbacks, it may be that this final option is the best that Peacocke can hope for, unless some more robust formulation of the Principal Hypothesis can be shored up.

**5. Explaining schizophrenic symptoms**

In the present section I shall consider one further argument that Peacocke offers in support of his Principal Hypothesis (interpreted as a claim about common underlying mechanisms), namely that it can explain the constellation of symptoms found in passivity forms of schizophrenia. It will emerge that, so far from providing support for his

view, a close examination of the data actually provides us with a further argument against it.

Schizophrenia comes in a number of different varieties. But people exhibiting so-called 'passivity symptoms' suffer from disorders of awareness of action, both physical and mental. Such people claim that their overt actions are not their own, even when the actions are appropriate to their conscious goals. A patient might say, for example (while acknowledging that he had decided to comb his hair), 'It isn't me who moves the comb through my hair, it is someone else.' These patients also suffer from auditory hallucinations, claiming that the voice that speaks inside their head (which is demonstrably their own inner speech) is really the voice of another person. And they claim, too, that *thoughts* are routinely inserted into their minds, often developing elaborate theories to explain how this happens (e.g., the CIA are somehow transmitting thoughts via radio waves). Peacocke argues that (key aspects of) this cluster of symptoms are neatly explained if we adopt his Principal Hypothesis, provided that the underlying damage is to the mechanism that generates efference copies of actions.[6]

One problem with this proposal is that these schizophrenic patients plainly *are* capable of creating efference copies from motor instructions, since they generate auditory imagery in the form of inner speech. For as we noted above, the best account of the creation of such imagery is that it begins from the efference copy of a mentally rehearsed action, which is transformed by an emulator system into a representation of what one might expect to hear if the action were executed. If Peacocke is to maintain that the problem in passivity-schizophrenia lies in efference copying, then he will need to claim that there are, normally, two sets of efference copies created. One of these would be used to generate appropriate imagery, and remains intact in schizophrenia, whereas the

6 Of course the Principle Hypothesis cannot provide a *full* explanation of these schizophrenic symptoms, as Peacocke points out. In particular, it does not explain patients' willingness to entertain bizarre explanations for the seeming heteronomy of their own actions.

other would be used to provide introspective access to the underlying action, and is damaged in schizophrenia. Such a claim would be motivated by nothing other than an attempt to save a theory.

Another problem with Peacocke's proposal, however, is that there is evidence that what has gone awry in passivity-schizophrenia is not the generation of efference copies as such, but rather that there has been damage to the comparator system that would normally take a forward model of the action and compare it with the incoming perceptual data (Frith *et al*., 2000a, 2000b). One strand in the evidence is the point that has just been made, that these patients are capable of generating visual and auditory imagery from mental rehearsals of action in something like the normal way. But in addition, they fail to show fast online repair of action and swift motor learning of the sort normally subserved by the comparator system. And perhaps most telling, one of the normal functions of the comparator system is to 'damp down' one's experiences whenever they match what would be predicted from the forward model. This is why it is impossible to tickle yourself (Weiskrantz *et al*., 1971; Blakemore *et al*., 1998). And it is also why you barely notice the sounds that you make when unwrapping a candy at the theatre while those around you are disturbed. But this 'damping down' effect fails to occur in passivity forms of schizophrenia. (Hence a schizophrenic patient *can* tickle himself.) It is this that is largely responsible for these patients' sense that their actions are not their own. For those actions are, quite literally, *experienced as if* the movements were being caused by another person.

Peacocke's Principal Hypothesis does not gain strength from its capacity to explain and unify the symptoms of passivity-schizophrenia, then. And in fact, the data provide us with a novel argument *against* his claims. Schizophrenic people report hearing voices, and they report generic thought insertion. But they never report *judgment* insertion, or *decision* insertion. Why not? If the root of the phenomenon were the lack of introspective access to mental actions (including, presumably, acts of judging and deciding), then this would be mysterious.

In the case of physical actions, Peacocke's claim is this. The patient knows on independent (observational) grounds that he is executing a certain action (combing his hair, say). But he lacks the normal introspective access to the motor schema that initiates and guides the action, resulting from damage to the efference copy system. Hence he claims that the action is not his own. Likewise, then, in the case of mental action. If the patient has independent evidence that he has just taken a decision (he has just reached out for one of two offered ice-creams, say), but lacks introspective access to the mental action of deciding, then one would expect him to report the situation by saying that *he* didn't choose the ice-cream: someone else chose for him. But this is not what he says. What he says is only that it is not he who *reaches for* the ice-cream. This is not what would be predicted if, as Peacocke claims, the mechanisms of self-awareness of both physical and mental action are the very same, and are damaged in passivity forms of schizophrenia.

In contrast, each of the two main competing views can explain these data. Those who maintain that there are separate self-monitoring mechanisms for mental and physical actions can say that these remain intact in passivity-schizophrenia, and that the problem is with the physical-action comparator system, just as Frith *et al*. (2000a, 2000b) claim. This will lead patients to claim that their physical actions are not their own, and neither are the mental episodes of inner speech that result from the mental rehearsal of some such actions. But their capacity to attribute mental acts of judging, deciding, and so forth to themselves will be unimpaired.

Those who claim that people's only access to their actions and attitudes is via interpretation of the behavioral and sensory (e.g., imagistic) data can likewise provide a successful explanation. For if the only relevant mental action that is introspectable by the agent is an act of saying something to oneself, then the facts are easily explicable. For only if an act of saying 'P' is *interpreted by the agent* as a judgment that P, or as a commitment to the truth of 'P', will the event in question have anything like the causal role of a judgment (Frankish, 2004). And it is just such an

interpretation that is absent in the case of a person suffering from passivity-schizophrenia, of course. Although he is well aware of the inner speech, a damaged comparator system leads him to deny that it is *his*. And in that case he will also deny that the speech-acts in question express his own commitments. So he will not be inclined to assert that anyone has inserted *judgments* or *decisions* into his head—just *words* (as well as thoughts in the generic sense of events with propositional content, of course).

**6. After Peacocke: the options**

I have argued that Peacocke's Principal Hypothesis is untenable. It is not the case that both mental and physical actions are available to subjects in exactly the same immediate non-perceptual way (and certainly not via the efference copies that those actions generate). This leaves open a pair of alternative accounts of the nature of self-knowledge, already briefly reviewed above. The first is that there are a number of distinct self-monitoring mechanisms, one of which might issue in non-perceptual knowledge of our own motor intentions and one or more of which might deliver non-perceptual knowledge of our own judgments, decisions, and so forth. This is the sort of account defended by Nichols and Stich (2003) and (in a somewhat different version) by Goldman (2006).

A second alternative would claim that non-inferential knowledge is restricted to our own perceptual and quasi-perceptual states, by virtue of their global broadcast and resulting availability as input to the mindreading system. Our access to our own actions and propositional attitudes, in contrast, is always via swift and unconscious interpretation of the available perceptual and contextual data. This is an account that has been defended, in various different guises, by a number of cognitive scientists (Gopnik, 1993; Gazzaniga, 1995, 2000; Wegner, 2002; Wilson, 2002), and it is, in my view, the one best supported by the psychological and neuropsychological data (Carruthers, 2009a). Such an account can explain the phenomenology of self-awareness just as well as Peacocke's, as we noted in Section 4. For we remain unconscious of the

interpretative process that issues in knowledge of our own actions and attitudes. Indeed, such an account would even justify a (very different) version of Peacocke's Principal Hypothesis, since it holds that one and the same mechanism (namely, the mindreading system, with access to perceptual and imagistic input) underlies awareness of both our physical actions and mental events of judging, deciding, and so forth.

Also left open, however, is the question of the extent to which, and in what ways, the mind is genuinely *active* in character. We noted in Section 3 that Peacocke has not done enough to establish that propositional attitude events of judging and deciding are really actions. For the fact that one can try, but fail, to make a judgment or reach a decision is just as consistent with the idea that judgments and decisions are the normal *effects* of actions, rather than *being* actions. But that does not mean, of course, that Peacocke's view is actually incorrect. Moreover, such a view could surely be rendered consistent with the self-monitoring accounts of self-knowledge described above. One could claim that judgings and decidings are indeed actions, of a sort, while claiming that the self-monitoring mechanisms that issue in non-inferential knowledge of them differ from those that are employed in awareness of our physical actions (while also perhaps differing from each other).

However, a starkly contrasting view of the place of action in the mind is also possible. And it is one that would comport quite nicely with the self-interpretational account of our knowledge of our own propositional attitudes sketched above. On this account, the only mental actions that exist are rehearsals of physical actions, which issue in inner speech and sequences of visual and other forms of imagery. These events often cause, or are caused by, judgments and decisions. But none of them *is* a judgment or a decision, and no judgments or decisions are themselves actions. Hence although actions, in the form of mental rehearsals, have an immense impact on the character and direction of our mental lives, our propositional attitudes remain intrinsically passive in nature. In closing, let me briefly elaborate this alternative conception of the active mind.

On this account, many of the distinctive properties of the human mind result from action rehearsals. Sometimes these are direct expressions (into speech, say) of judgments, decisions, or other attitudes. But often they are undertaken for purposes of their own—sometimes for pleasure, as in fantasizing, sometimes for problem solving, as when one tries out candidate actions in imagination, responding emotionally to the scenarios that we construct. Sequences of action rehearsal can take place in ways that are habitual, or they can be guided by our beliefs—especially by our normative beliefs about the ways in which one *should* think and reason (either in general, or in specific situations). The result is what Dennett (1991) calls the 'Joycean machine', or what cognitive scientists refer to as 'System 2' reasoning (Evans and Over, 1996; Sloman, 1996, 2002; Stanovich, 1999; Kahneman, 2003; Carruthers, 2009b). And these sequences of mental rehearsal can give rise to novel attitudes, and to novel physical behavior. But none of the various instances of action-rehearsal *is* an attitude. For example, at the conclusion of an episode of reflective System 2 reasoning I might say to myself in inner speech, 'So, I won't go to the pool this afternoon.' Although this has the superficial form of a decision, it is not one. But as a result of it, I am apt to interpret myself as having decided not to go to the pool, or as having committed myself not to go (Frankish, 2004). And this is likely to have consequences for my subsequent behavior (whether overt or covert) that are very much *like* the consequences of a decision, as we noted earlier.

On both accounts, then—Peacocke's and the one just sketched—the human mind is thoroughly infused with, and partly constituted by, mental actions. But the accounts differ profoundly over the question whether propositional attitude events (judging, deciding, and the rest) are themselves actions, or are merely influenced by actions. Attempting to adjudicate between these competing conceptions of the active mind is too large a task to be undertaken here. The issues are complex, and many different bodies of data are relevant, and would need to be

considered. But it is worth noting that in reaching a negative evaluation of Peacocke's Principal Hypothesis we have left open large and deep issues about the place of action within the human mind.[7]

*University of Maryland, College Park*
*pcarruth@umd.edu*

## References

Baars, B. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Science*, 6, 47-52.

Baars, B., Ramsoy, T., and Laureys, S. (2003). Brain, consciousness, and the observing self. *Trends in Neurosciences*, 26, 671-675.

Blakemore, S., Wolpert, D., and Frith, C. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, 1, 635-640.

Carruthers, P. (2009a). How we know our own minds: the relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, 32 (2), 121-138.

Carruthers, P. (2009b). An architecture for dual reasoning. In J. Evans and K. Frankish (eds.), *In Two Minds*, Oxford University Press.

Carruthers, P. (2010). Introspection: divided and partly eliminated. *Philosophy and Phenomenological Research*, 79.

Dennett, D. (1991). *Consciousness Explained*. Penguin.

Dehaene, S. and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, 79, 1-37.

Dehaene, S., Sergent, C., and Changeux, J. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Science*, 100, 8520-8525.

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

7 I am grateful to Fabian Dorsch, Brendan Ritchie, and an anonymous reviewer for this journal for their insightful comments on an earlier draft.

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

Frith, C., Blakemore, S., and Wolpert, D. (2000a). Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London*, B, 355, 1771-1788.

Frith, C., Blakemore, S., and Wolpert, D. (2000b). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Research Reviews*, 31, 357-363.

Ganis, G., Keenan, J., Kosslyn, S., and Pascual-Leone, A. (2000). Transcranian magnetic stimulation of primary motor cortex affects mental rotation. *Cerebral Cortex*, 10, 175-180.

Gazzaniga, M. (1995). Consciousness and the cerebral hemispheres. In M. Gazzaniga (ed.), *The Cognitive Neurosciences*, MIT Press.

Gazzaniga, M. (2000). Cerebral specialization and inter-hemispheric communication: does the corpus callosum enable the human condition? *Brain*, 123, 1293-1326.

Goldman, A. (2006). *Simulating Minds*. Oxford University Press.

Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27, 377-442.

Jeannerod, M. (2006). *Motor Cognition*. Oxford University Press.

Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist*, 58, 697-720.

Kosslyn, S. (1994). *Image and Brain*. MIT Press.

Kosslyn, S., Thompson, W., Wraga, M., and Alpert, N. (2001). Imagining rotation by endogenous versus exogenous forces: distinct neural mechanisms. *NeuroReport*, 12, 2519-2525.

Lamm, C., Windtschberger, C., Leodolter, U., Moser, E., and Bauer, H. (2001). Evidence for premotor cortex activity during dynamic visuospatial imagery from single trial functional magnetic resonance imaging and event-related slow cortical potentials. *Neuroimage*, 14, 268-283.

Nichols, S. and Stich, S. (2003). *Mindreading*. Oxford University Press.

Paulescu, E., Frith, D., and Frackowiak, R. (1993). The neural correlates

of the verbal component of working memory. *Nature*, 362, 342-345.

Peacocke, C. (2007). Mental action and self-awareness (I). In J. Cohen and B. McLaughlin (eds.), *Contemporary Debates in the Philosophy of Mind*, Blackwell.

Peacocke, C. (2008a). Mental action and self-awareness (II): epistemology. In L. O'Brien and M. Soteriou (eds.), *Mental Action*, Oxford University Press.

Peacocke, C. (2008b). *Truly Understood*. Oxford University Press.

Prinz, J. (2007). All consciousness is perceptual. In J. Cohen and B. McLaughlin (eds.), *Contemporary Debates in the Philosophy of Mind*, Blackwell.

Richter, W., Somorjat, R., Summers, R., Jarnasz, N., Menon, R., Gati, J., Georgopoulos, A., Tegeler, C., Ugerbil, K., and Kim, S. (2000). Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *Journal of Cognitive Neuroscience*, 12, 310-320.

Shergill, S., Brammer, M., Fukuda, R., Bullmore, E., Amaro, E., Murray, R., and McGuire, P. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping*, 16, 219-27.

Sloman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.

Sloman, S. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin, and D. Kahneman (eds.), *Heuristics and Biases*. Cambridge University Press.

Stanovich, K. (1999). *Who is Rational?* Lawrence Erlbaum.

Weiskrantz, L., Elliot, J., and Darlington, C. (1971). Preliminary observations of tickling oneself. *Nature*, 230, 598-599.

Wegner, D. (2002). *The Illusion of Conscious Will*. MIT Press.

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

Wolpert, D. and Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, 3, 1212-1217.

Wolpert, D. and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11, 1317-1329.