Chapter 12

Can neuroscience explain consciousness?

(Originally published in *Journal of Consciousness Studies*, 11 (7-8): 180-198, 2004)

(with Chris Frith)

Preview

Cognitive neuroscience aspires to explain how the brain produces conscious states. Many people think this aspiration is threatened by the subjective nature of introspective reports, as well as by certain philosophical arguments. We propose that good neuroscientific explanations of conscious states can consolidate an interpretation of introspective reports, in spite of their subjective nature. This is because the relative quality of explanations can be evaluated on independent, methodological grounds. To illustrate, we review studies that suggest that aspects of the feeling of being in control of one's bodily movement can be explained in terms of the complex and surprising way the brain predicts movement. This is a modest type of functional, contrastive explanation. Though we do not refute the threatening philosophical arguments, we show that they do not apply to this type of explanation.

Chapter 12

Can neuroscience explain consciousness?[1]

(with Chris Frith)

1. Introduction

There is very widespread agreement that, on the one hand, there has been considerable progress in identifying the neural correlates of consciousness (NCC), but that, on the other hand, there has been next to no progress in using the NCC to explain how brain activity gives rise to conscious states (Kanwisher 2001, Laureys et al 2000; Block 2001, Chalmers 2000; Noë and Thompson, 2004, Hohwy and Frith 2004; Portas et al 2003, Block and Stalnaker 1999: 1, Machamer et al 2001: 4). Many people think that the reason for this impasse is the subjective nature of conscious states. Some philosophers argue that objective neuroscience *cannot* capture the subjective nature of what it is like to have conscious states (e.g., Levine 1983, Jackson 1986, Chalmers 1996), whereas some neuroscientists are wary of the subjective nature of the introspective reports that are the raw data for the study of consciousness. Some of these neuroscientists try to exorcise the use of introspective reports as much as possible, while others argue that better methods for interpreting the reports is one of the keys to explaining consciousness (e.g., Jack and Shallice 2001, Jack and Roepstorff 2003).

We think that the reason for the current lack of explanation may be due more to the fact that consciousness is fiendishly difficult to explain than to the subjectivity of consciousness. We argue

[1] Earlier versions of this paper was presented at philosophy seminars at Monash University and Adelaide University, and at a seminar at the Centre for Functionally Integrative Neuroscience, Aarhus University Hospital. We wish to thank the participants for comments and suggestions. The research for this paper was supported by the Danish Research Council for the Humanities and the Wellcome Trust.

that some of the wariness about introspective reports may subside once explanations become available. We illustrate this with an example where neuroscience does in fact explain, at least in part, how brain activity gives rise to a conscious state: aspects of the conscious feeling of being in control of one's bodily movements is explained by a neural mechanism that predicts those movements in a certain way. This is challenged by the philosophers' sceptical arguments and, though we do not refute these arguments, we go on to argue that they do not apply to the type of explanation in question.

2. Explanation and the interpretation of introspective reports.

Some of the wariness about accepting introspective reports probably has something to do with the protracted effects of the behaviourist attack on the introspectionism of the late 1800s and the early 1900s (see, e.g., Baars 2003). Here we focus on a recent, more measured form of caution about introspection. Jack and Shallice usefully put it in terms of the interpretation of introspective reports:

> It is hazardous to place any reliance on generalizations derived from experience unless they can be validated by objective evidence. This is because the principal problem with the use of introspective evidence is that it is prone to misinterpretation (2001: 176).

Misinterpretation can arise in two ways: as a result of flaws in the conceptual framework employed by the subjects themselves in their attempt to express their conscious states, and/or as a result of flaws in the experimenter's conceptual framework for interpreting the subject's reports. It is very difficult to devise experimental paradigms and conceptual frameworks that can exclude the possibility that subjects either use different words to express the same feeling, or the same words to

express different feelings (for a recent discussion, see Schooler 2002). The effect of misinterpretation is to introduce damaging inaccuracy in the raw data for the study of consciousness.

We shall exploit the following intuition. If there were a good neuroscientific explanation of a type of conscious state, then that would suggest that the introspective reports in the studies underlying that explanation were correctly interpreted; it seems unlikely that a good explanation would be available, if the interpretations of the reports had been very inaccurate.

The problem with this intuition, quite apart from the apparent lack of explanations already noted, is that it can easily create a vicious circle. On the one hand, the introspective reports would constitute an essential part of the evidence for the explanation in question. On the other hand, the explanation would in part make us understand why those reports occurred. If there are doubts about the nature of the reports—e.g., worries about misinterpretation—then we cannot use the explanation to validate the interpretation of the reports because the reports are themselves part of the reason we believe the explanation (Hempel 1965; Lipton 1991). We should not have believed the explanation in the first place if the reports are not to be trusted.

However, on an alternative conception of the problem it is possible to break out of this circle. Assume that the same set of introspective reports is interpreted by one experimenter using the conceptual framework *C1*, and is interpreted in an incompatible way by another experimenter using the framework *C2*. At least one of them must be misinterpreting the subjects. No new evidence, in the form of more introspective reports, can decide which framework is the appropriate one because all such new evidence will be interpreted according to the preferred framework of each

experimenter. Further behavioural and neurophysiological data can aid interpretation (as discussed by Jack and Shallice 2001), but the subjective and private character of consciousness may often appear to put a solution to the conflict between *C1* and *C2* out of reach.

It is a sound principle of science that we should accept theoretical and conceptual frameworks that help provide better explanations over frameworks that provide less good explanations. If *C1* is associated with better explanations than *C2*, then that would give us reason to prefer the *C1*-interpretations of the introspective reports. This would not involve a vicious circle because our reason for accepting *C1* would have to do with independent general, methodological considerations about what makes an explanation *better* (here we adapt an argument in Lipton's discussion of the notion of inference to the best explanation, 1991: 70–71; this general kind of strategy is further explored in Hohwy 2004).

We have set out a non-circular strategy for consolidating a given interpretation of introspective reports. On this strategy, it is not the case that we first need to decide which interpretation of the data is correct, and only then begin to formulate explanations. Rather, good interpretations and good explanations evolve in tandem. To support this strategy we now present a type of conscious state and provide a detailed example of a good, if partial, neuroscientific explanation of a conscious state.

3. The conscious feeling of being in control of one's bodily movements.

The case we present stems from studies in cognitive neuroscience of the feeling of being in control of one's bodily movements. The question concerning the right interpretation of subjects' reports of feeling in control, and of their various reports when they successfully and voluntarily lift their arm, is pertinent because the feeling of being in control has a rather elusive, 'thin' phenomenology

(Haggard and Johnson 2003); it may belong, in James' terms, to fringe consciousness (James 1890; see also Baars 2003). It can therefore be difficult to distinguish the feeling of being in control from among the cluster of feelings concerning different types of willed action.

In table 1, we contrast the feeling of being in control with three other feelings: the feeling of passive movement, the feeling of not being in control, and the feeling of being the agent behind an action. This reveals our conceptual framework for interpreting introspective reports in this area of study.

Type of conscious state	Typical conditions	Example	Representative study	Subjective report
Feeling of being in control of bodily movement	Successful, intentional movement of limbs	You want to lift your arm and succeed in lifting it	Blakemore et al 2003.	Rate the 'voluntariness' of the movement on a scale from 0 (voluntary) to 100 (involuntary).
Feeling of passive movement	Having one's limbs moved in the absence of intention to move	Someone else lifts your arm	Blakemore et al 2003	Rate the 'voluntariness' of the movement on a scale from 0 (voluntary) to 100 (involuntary).
Feeling of not being in control	Intentional movement of one's limbs where the immediately desired state is not achieved.	You reach for an object as it unexpectedly moves You experience very distorted feedback of your movement	Castiello et al 1991 Franck et al., 2001	Report when you saw the target move Report whether the feedback corresponded exactly to the movement
Feeling of being the agent of an action	Performing an intentional action	You manipulate a computer mouse and stop it at a certain spot on the screen	Wegner and Wheatley 1999	Rate how much you intended to make the stop (scale with percentage marks 0-100 between 'I allowed the stop to happen' and 'I intended to make the stop'.

Table 1

This cluster of feelings is also related to, but different from, the feeling of intention, or 'conation', that is, a conscious mental state aiming at producing a specific goal through a physical action (Haggard and Johnson 2003; Haggard 2001). This feeling is a conscious intentional state that often precedes voluntary action (though probably not when action is automatized).

Notice that we do not mean to imply that these feelings are mutually exclusive: intuitively, it seems that the feeling of being the agent behind an action can accompany the feeling of being in control (and, in rarer cases, it may accompany the feeling of not being in control of one's bodily movements). Though the movement of limbs is often the means to perform some action (e.g., manipulate an object), limb movement may sometimes be an action in itself (e.g., a ballet dancer rehearsing new steps). So often the feeling of being the agent and the feeling of being in control will both be part of fringe consciousness.

James Russell (1996: 78–79) gives a vivid description of the feeling of being in control, and in turn anticipates aspects of the explanation we present below.

> [I]n a pitch-black room wearing a suit to insulate you from all tactile inputs you will know you have raised your arm or stood up without needing to look at your body and without feeling it brush against something: kinaesthetic feedback will be sufficient. But my point is not merely about kinaesthesia. You will know that you have acted because you will have launched the motor command and because you will have made predictions about the phenomenal outcome of your launched movement. In that sense you know "immediately" that you have acted; you do not have to wait for the world to tell you whether you have done so.

4. From neural correlate to explanation.

Being in control and the attenuation of sensation

A number of studies suggest that the feeling of being in control correlates with a reduction of activity in inferior parietal cortex, often including the secondary somatosensory cortex (SII) in the parietal operculum. This area is normally more active in passive arm movements and when the subject perceives or imagines the movements of other people (Sirigu et al 1999, Blakemore et al 1998, Weiller et al 1996, Decety et al 2002, Farrer and Frith 2000). Similarly, in delusions of alien control in schizophrenia there is overactivity in this area (Spence et al 1997). In one important study, Blakemore et al (2003) isolated the feeling of being in control from the performance of active movements. Susceptible volunteers were hypnotised and scanned in three conditions: while their arm was passively moved up and down, while they actively moved it, and while they actively moved it but were given the hypnotic suggestion that their arm was being moved passively. In the third condition the volunteers lost the feeling of being in control and felt that their arm was being passively moved up and down. Compared with the active movement condition, there was greater activity in the inferior parietal cortex and the cerebellum during passive movement when the volunteers didn't feel in control *and* when the movements were active but weren't associated with a feeling of being in control, so it seems that activity in this area is specifically associated with the feeling of being in control. These studies similarly suggest that the inferior parietal cortex is involved in our important ability to distinguish between self-produced actions and actions produced by others (for a review see Meltzoff & Decety, 2003). There is some evidence that the anterior cingulate and, in turn, the dorsolateral prefrontal cortex play a role in modulating the activity in the inferior parietal cortex (Spence et al 1997, Fletcher et al 1999).

Theoretical and experimental work on the motor control system enable a better understanding of the constitutive role of the neural correlates of the feeling of being in control. Before a subject intentionally moves her arm, a motor command is issued. The motor command is arrived at via an inverse model that begins with the desired state and calculates which movement could be performed to achieve it, given the body's current state. When the motor command is issued so is an efference copy in the form of a forward model that allows the accurate prediction of two things: (i) how the current state will change given the motor command; (ii) what the sensory feedback will be given the new state (Wolpert and Ghahramani 2000). There is evidence that the cerebellum is involved in predicting the sensory feedback of self-produced movements (Wolpert et al 1998) and that the forward model is stored in the lateral cerebellar cortex (Imamizu et al 2000). The same region is differentially activated according to the specific consequences of particular movements (Blakemore et al 1998, 1999b), and its activity increases as the actual consequences of those movements deviates from the predicted sensory consequences (Blakemore et al 2001).

It is currently believed that activity in the inferior parietal cortex is modulated by the forward model's prediction of the sensory consequences of movement. Thus, activity in the inferior parietal cortex is attenuated in anticipation of that sensory input. In contrast, when there is no self-initiated movement, there is no forward model and no attenuation of activity. The mechanism of forward modelling and attenuation provides the mechanism that explains how the feeling of being in control arises: The forward model predicts what movement is about to happen and this in turn ensures that the sensory consequences of the movement, though registered by the brain, do not acquire informational salience to the subject, as this is not something the subject needs to pay attention to. As Portas et al say in their review of the literature 'the feeling of being in control arises precisely

when all our predictions and expectations about our movements are fulfilled' (Portas et al 2003, 291). It is important to notice that the notion of 'prediction' used here is different from the more usual notion of conscious prediction (as in, e.g., I predict that it will rain). The computations of the forward model happen at a subpersonal level, and we may only be consciously aware of the resulting predicted state (Blakemore et al 2002: box 1).

The proposal is not that this mechanism *alone* explains *everything* about how the conscious feeling of being in control arises. Obviously, many other elements will eventually have to go into that explanation. Instead, the proposal is that this mechanism explains why it feels like the feeling of being in control, rather than like other closely related feelings. Thus, consider the contrast to the feeling of passive movement (e.g., when your arm is lifted by someone else). Such events are associated with more activity in the inferior parietal cortex because no forward model is produced and there is no corresponding attenuation. The subject will have the feeling of being passively moved partly because the sensory consequences of the movement will be informationally salient.

Being in control and the on-line adjustment of movement

The explanation is complemented by consideration of a further contrast. Why does the forward modelling and attenuation mechanism give rise to the feeling of being in control, rather than just the feeling of having predicted something (I can sometimes predict the rain, but I don't feel in control of the rain)? Based on the inverse model's mapping of motor commands for each point along the trajectory of the movement, the forward model also allows prediction of how the actual state will change. These predictions can be compared with the desired outcome, and the comparisons can be used to make fine adjustments to the ongoing motor commands before reafferent feedback from the movement is available (Desmurget & Grafton 2000). This happens preconsciously (Blakemore et al

2002) and makes rapid error correction possible. A hypothesis about certain conditions of schizophrenia suggests that people with schizophrenia may have a malfunction with the system of forward modelling (Frith 1992, Frith et al 2000), and indeed this hypothesis explains the observation that they are relatively poor at rapid error correction (Frith and Done 1989). The special role of the forward model therefore allows the subject to ignore the need for fine adjustments to her ongoing movements because this modelling pre-empts the source of data (reafferent feedback) one would normally think is needed for such adjustments. When those data come in, the adjustments have already been made. This may explain why the subject's ability to predict what is going to happen results in more than just the feeling of having predicted it.

Core components of the explanation

Thus our explanation of the feeling of being in control has two core components: the prediction and attenuation mechanism, and the comparison and adjustment mechanism. Both are necessary for the explanation to succeed. If one could make adjustments to one's movements, but had to pay attention to them, then one would hardly feel in control in the way we normally do. If one could afford not to pay attention to one's movements, but couldn't make adjustments to them, then one would hardly feel in control in the way we normally do. The demonstrations of sensory attenuation and of automatic on-line adjustments are derived from experimental studies of motor control the results of which stand whether or not we consider their relation to conscious experience. We now elaborate the explanation by considering further contrasts.

Not feeling in control: When the sensory consequences of our movements are important

Consider now the contrast to the feeling of not being in control. It is part of this story that when the discrepancy between the intended and the actual outcome is large, the actual sensory

consequences often (but not always, Fourneret and Jeannerod 1998) become accessible to awareness (Franck et al., 2001; Blakemore et al 2002). Presumably, when it becomes unlikely that we can achieve the intended outcome relatively easily, or when it seems likely that another agent is in play whose intentions we may not know, we need to attend directly to the situation. This is a situation where the feeling of being in control momentarily disappears because the subject can't any longer afford to ignore the sensory consequences of her movements and she may no longer be able to achieve the desired goal by relying on preconscious fine adjustments. The subject will normally regain the feeling of being in control, when the new adjustments have been made, because those adjustments again rest on issuing motor commands, with their associated forward models and attenuation.

A variant of this case is when there is no discrepancy between the intended and the actual outcome except that the intended outcome (here, a tickle produced via a mechanical device) comes later than predicted by the forward model (Blakemore et al 1999a). In these cases, the subjects are not aware of the discrepancy, but they have an increased rating of the perceived tickliness, no doubt because the delay is not predicted and there is less attenuation (the introspective report in this kind of study is solicited by asking subjects to rate how intense, pleasant and ticklish stimulation of the palm is). To our knowledge, this kind of experiment has not been made of bodily movement, but we would predict that delayed arm movement would be accompanied by a diminished feeling of being in control.

Not feeling in control: When movement requires concentration

There is the case of Ian Waterman, described by Jonathan Cole (1991), who has lost all sensation (except heat, pain and deep pressure) from the neck down. He can't feel where his arms and legs

are. This deafferented condition usually leads to immobility even though motor output is intact. Immediately after losing sensation, IW had no control over his attempted movements, and his arms would move involuntarily when out of his sight. IW has since taught himself to move quite well, but must constantly attend to visual feedback. We have been arguing that the feeling of being in control does not depend as much on reafferent feedback as on predicting the reafferent feedback, so IW could perhaps have retained some feeling of being in control. However, two things may complicate matters. First, the conscious and strenuous effort that IW puts into moving may alter the way he feels when he moves. We have suggested that part our normal feeling of being in control relates to our ability to make automatic, on-line corrections to our movements. This aspect of the feeling of being in control is presumably not experienced by IW. Second, on the model of motor control discussed above, the absence of somatosensory feedback would interfere with the normal course of forward modelling and prediction of movement. This is because prediction requires state variables, that is, information about the state of the body prior to implementation of the motor commands, in addition to copies of motor commands. Otherwise one couldn't predict the starting and ending points of the movement. IW can only obtain information about state variables through vision, not through the normal routes. So there is reason to think that his motor control system doesn't operate in the way we have described above (for further discussion, see, e.g., Stenneken et al 2002, see also Fourneret et al 2002 who consider the role of comparisons between forward models and reafferent sensory signals for conscious recognition of action). Though we know of no specific report of the extent to which IW *feels* in control, it seems likely that he doesn't feel completely out of control (except when deprived of visual feedback in which case he falls to the ground), but that he consciously and with a lot of mental effort has to compute his actual state and all adjustments to his ongoing movements (Cole 1991).

The feeling of agency

Consider now the contrast to the closely related feeling of being the agent behind an action. A natural thought is that subjects feel in control precisely when they feel they are the agents behind an action. So why is the feeling of being in control related to forward modelling and attenuation, rather than the feeling of agency? We think the two feelings are distinct, but that they often, though not always, come together (to illustrate, a soccer player who hurls himself into a cluster of players, trips and fortuitously scores a goal may plausibly feel like the agent even though he feels he is not in control of his bodily movements). To clarify, we briefly discuss an influential study of the feeling of agency by Wegner and Wheatley (1999) in which subjects believed that they were the agents behind certain events even though they did not have the corresponding intentions, and even though a confederate was the agent behind the event. They apparently judged this on the basis of the temporal contiguity between a conscious thought about the event and the occurrence of the event. Thus, a participant and a confederate manipulated a computer mouse together for about 30 secs at a time while listening to instructions over head sets. The participant would be primed with words for objects, some of which were seen on the computer screen's picture of a number of little objects. It was first established that priming the participant did not cause him or her to move the cursor any closer to the mentioned object than when the object was not mentioned. Next, the confederate would on some trials be instructed to unobtrusively make the cursor stop at a particular object and the participant would be primed with this object either 30, 5 or 1 second before the stop, or 1 second after the stop. In this task the participant's thought about the object, induced by the priming, was not causally efficacious in making the cursor stop at that object. Nevertheless, the participants reported an increased experience of intention for stops when they had been primed 1–5 secs earlier (mean percentage rated intentionality of about 60%). The experience of

intention was rated as considerably lower when participants were primed 30 secs earlier or 1 sec after (mean percentage rated intentionality of about 45%).

Wegner & Wheatley suggest a three-factor model for self-attributions of agency: one's thoughts about the action need to be (1) temporally prior to the action, (2) consistent with it, and (3) it should be the only [known] cause of the action. Even though reliable contingency allows prediction, this model has no obvious need for the story about forward models as predictions of the motor control system. The sense of agency simply arises when the subject experiences the action and the thought about the action as satisfying those three conditions. Since there seems to be a close relation between the sense of agency and the sense of being in control, this model could spell problems for our explanation. Couldn't the model of Wegner & Wheatley explain the feeling of being in control better?

However, the model seems insufficient because in certain cases all three conditions appear to be satisfied and yet there is no feeling of being in control. This happens in schizophrenia, for people with delusions of control who intend to perform an action, are aware of their intention, perform the action successfully, and yet attribute initiation of the action to someone else (Spence et al 1997). This interpretation of reports of the delusional experience is based on a comparison of the behaviour and introspective reports between subjects with anarchic hand sign and subjects with delusions of control, interpreted within the general framework of motor control described above (Frith et al 2000a, 2000b). Subjects with anarchic hand sign say things like 'it will not do what I want it to do' (Goldberg et al 1981, quoted in Frith et al 2000a). Subjects with delusions of control say things like 'The force moved my lips. I began to speak. The words were made for me' (Mellors 1970, quoted in Frith et al 2000a), and 'I felt like an

automaton, guided by a female spirit who had entered me during it [an arm movement]' (Spence et al 1997). For an outside observer there may be nothing untoward about the movement of the deluded patient, in contrast to the subject with anarchic hand sign; and in the case of delusions of control the actions made while the patient feels she is being controlled are consistent with her intentions (they may correspond to the instructions the patient has just been given by the experimenter). Moreover, delusions of control can be explained within the forward modelling framework, in terms of a malfunction with the issuing of forward models that results in false discrepancies between intended and actual outcomes (Frith 1992, Frith et al 2000b): the result is that the patient intends to perform an action and performs it, but experiences it the way she experiences passive movement. This suggests that the explanation of the feeling of being in control in terms of forward modelling and attenuation does have a distinct role to play, possibly alongside Wegner and Wheatley's account of the feeling of agency.

Timing in the awareness of action

Studies of the feeling of being the agent of an action show that it depends on the temporal relationship between intention and action, and action and consequence, and the perceived times of these events (e.g. Haggard et al. 2002). Some of these studies relate to the study of the feeling of being in control, and the findings integrate well with the proposed mechanisms that give rise to that feeling.

In the experiment of Libet et al. (1983), subjects were instructed to lift a finger 'whenever they felt the "urge" or desire to do so'. Subjects were also asked to indicate the exact time at which this 'urge' occurred. This was found to be at about 200-300 ms before the actual start of the

movement. If EEG measurements are made during spontaneous movements of this type then gradual increases in negativity can be seen for up to 1000 ms prior to the onset of the movement (the readiness potential). Libet et al. also measured the readiness potential in their study and showed that the awareness of the urge to move occurred after onset of the readiness potential, potentially posing a challenge to the notion that our conscious intentions are the causes of our actions.

The readiness potential consists of a number of components. Initially the signal is bilaterally symmetric, but at about 500 ms prior to the movement it becomes asymmetric with larger amplitudes over the hemisphere contralateral to (and controlling) the moving hand. Presumably the onset of the lateralised component of the readiness potential corresponds to the time at which the response to be made has been specified at the neural level. Haggard & Eimer (1999) found that the time of the awareness of the 'urge' to move correlated with the time of the onset of the lateralised component of the readiness potential, rather than earlier stages. This observation suggests that the awareness of the 'urge' to act is linked to the neural process that specifies the act to be made.

In contrast to awareness of the urge, awareness of initiating the act (e.g., when one's finger begins to move) precedes the onset of the actual movement by 50-80 ms in the Libet paradigm (Libet et al 1983, McCloskey et al 1983, also Libet 1985, Haggard and Magno 1999). Therefore awareness of initiation cannot depend on sensory feedback from the moving limb, since this does not reach the brain until many ms after the onset of the movement. Instead it may depend on the predicted sensory consequences of the limb derived from the forward model. Awareness of having the urge to act thus requires specification of the movement,

whereas awareness of having initiated the act must await the computation of the sensory consequences of the act (Frith et al 2000a).

This has the effect that when the sensory feedback then begins to come in, the subject is already aware of having initiated the movement on the basis of the predictions of the forward model. This integrates well with the explanation of the feeling of being in control because it would not be plausible that the subject on the one hand could ignore the sensory consequences of the movement and thus feel in control, and on the other hand have to pay notice to the sensory consequences in order to be aware of having initiated the action.

An interesting feature of the Libet paradigm is that cause (the urge to act) and effect (initiating the act) are closer together in awareness than they are in terms of underlying physical processes (onset of lateralised readiness potential and onset of movement). Such an attraction effect is also observed when a voluntary action (e.g. pressing a button) is followed by some external consequence (e.g., a tone). There is an attraction effect (called intentional binding) in that the action and its consequence are perceived to be closer together in time than they really are (interestingly, intentional binding is stronger in patients with schizophrenia than in normal controls, Haggard et al 2003). The effect is reversed when the action is involuntary (such as a muscle twitch induced by transcranial magnetic stimulation), and is modulated by temporal contiguity and predictability (Haggard et al 2002). The time of the involuntary act and its consequence are perceived to be further apart than they really are.

Intentional binding is relevant for the sense of being the agent of an action. It perceptually binds the actions we are aware of initiating, and that we feel in control of, to their predicted, contiguous

consequences. This may be what constitutes our sense of agency (Frith 2002b), and is in contrast to what happens when we are not aware of initiating the action, and when we do not feel in control. It may be that the mechanism that underlies the feelings of being in control and the awareness of initiating the action play a role in facilitating the perception of temporal contiguity, and thus of agency, in the cases where we really are the agents.

5. Explanation characteristics

Among philosophers of science there is no real agreement on a set of necessary and sufficient conditions for something to be an explanation (Newton-Smith 2000: 132). So there is no straightforward and uncontroversial way to defend the claim that we have provided an explanation of a conscious state. We trust that the above account helps making this claim plausible and below we support the claim by discussing some of the theoretical characteristics of the explanation.

What makes the explanation best?

We have discussed the feeling of being in control by contrasting it with other feelings (of passivity, of feeling not to be in control, of having predicted something, and of agency). One question is thus whether the mechanism concerning forward modelling and attenuation would do a better job at explaining one of these other feelings, than of explaining the feeling of being in control. In each case we find that the mechanism explains the feeling of being in control best. The other candidate explanations would integrate poorly with the overall picture of types of conscious states (cf Table 1) and/or with neurophysiological findings (e.g., the Libet studies, Blakemore et al 1998, Blakemore et al 2003), and/or findings from psychopathology (e.g., rapid error correction in schizophrenia). Another question is whether some other mechanism (i.e. that associated with Wegner and Wheatley 1999) could explain the feeling of being in control but again we found that this candidate

explanation, though perhaps appropriate for the feeling of agency, doesn't integrate well with psychopathological cases (delusions of control).

It is a sound principle of science that one should give most credence to the better of two rivalling explanations, and that overall integration is an important independent criteria for deciding which explanation is better (Bird 1998: Ch 2). It follows that we have reason to accept the forward modelling and attenuation account, including—as we promised at the beginning—its associated interpretation of the introspective reports in question. (Notice that this is far from *guaranteeing* the truth of the account and the infallibility of the interpretations since it is, as always in science, possible that later discoveries and theories will throw up even better explanations).

It is worth noticing that the explanation also integrates well with theories on a larger scale. Thus the forward modelling motor control theory builds on well-established theories in engineering (Wolpert et al 1995), and there are suggestions that forward modelling mechanisms may play a role in other representational systems than the system devoted to self-initiated and externally initiated movement, such as representations of the mental states of other agents (Wolpert et al 2003), and in visual perception (Friston 2002, Eliasmith 2003, Grush, forthcoming).

What type of explanation?

The explanation of the feeling of being in control fits well with one influential notion of explanation: contrastive explanation (Lewis 1986, Lipton 1991, 1993; related to Mill's method of difference, Mill 1904: 3.8.2; see also Baars 1997). This is because the explanation is naturally represented as progressing by answering a series of contrastive questions of the type "why *X*, rather than *Y*?" For example, "why is this mechanism related to the feeling of being in

control rather than merely to the feeling of having predicted what will happen?" In relation to questions of constitution, contrastive explanations serve to describe something about the constitution of a certain entity that accounts for the highlighted contrast (for example, we might explain why one metal, rather than another, developed metal fatigue in the same conditions by citing some relevant facts in the constitution of one, that are not found in the other).

In science and in everyday life we very rarely attempt to explain things completely and from the bottom up—we would not know where to begin or where to stop. Instead, we have some relevant contrast in mind that serves to pick out an aspect we, given our epistemic standpoint, are particularly interested in. This strategy often eases the explanatory burden considerably. This also has the effect that contrastive explanations are very selective: only certain aspects of the phenomenon we are interested in will get explained. Hence, the above explanation is not intended as a complete explanation of the feeling of being in control (let alone of consciousness); it is intended as a complete explanation of a few aspects of the feeling of being in control, as highlighted through contrasts with some other feelings.

What is doing the explaining?

The forward modelling and attenuation account is basically a cognitive explanation that in principle can be told without recourse to a description of the neural correlate of consciousness: the explanatory work is to a large extent being done by the cognitive and computational mechanisms, and these mechanisms are then simply implemented in neural matter (we could imagine it being implemented in some other matter) (for discussion of the notion of mechanism, see Cummins 2000, Craver and Darden 2001, Machamer et al 2000). So in *principle*, for this type of explanation, there

is no need for neuroscience. But in *practice*, the study of neural activity is inescapable. Firstly, study of the neural correlates of conscious states is needed to ensure that the proposed cognitive mechanisms are indeed explanatory of a particular conscious state (and not of another conscious state or of a mere belief state). Secondly, neuropsychological studies, as well as detailed studies of neurophysiology and neurochemistry, are needed (a) in a kind of heuristic role to suggest which cognitive mechanisms may be in play and (b) to subsequently test whether the proposed cognitive mechanisms are indeed implemented by the identified brain structures (see also Frith 1992: Ch. 3)

Though the explanation is at heart functional, the mechanism involving forward modelling and attenuation is not something we would expect on the basis of mere a priori, theoretical reflection on the concepts involved or reflection on the merely discriminatory abilities of the subjects. Prior to doing the empirical studies, there seems no reason that predicted events are accompanied by the feeling of being in control since we sometimes do things that are not predicted even though we feel in control as we do it (for example, being able to catch a bottle of wine just about to topple over), and often we do not feel in control of events that we have predicted. It is only when the neurophysiological evidence comes in that we see how predictions, in the form of subpersonal forward models that attenuate certain activity, can be related to the feeling of being in control. There is also no a priori reason to think that the feeling of being in control is intimately connected with a pattern of prediction of the sensory consequences of one's bodily movements, rather than with a more direct, on-line control mechanism. Lastly, there seems no a priori reason to say that the feeling of being in control should be explained by a *reduction* of activity, that is, that the feeling of being in control arises because there are things we are made to ignore. It is like the mirror image of what we should

expect a priori: the feeling of being in control is explained by our not having to control what happens.

What makes it explanatory?

This is a one of the hardest questions to answer. If there were a straightforward answer, then there would be a lot less controversy in philosophy of science about the correct analysis of the concept of explanation. In our case it seems to help that we discuss the feeling of being in control together with a cluster of other, more or less well studied, feelings of the same volitional family; and that we use the neurally implemented mechanism to distinguish the feeling of being on control from those other feelings with which it shares some phenomenal aspects (for some discussion of this type of account, see Clark 2000). So the explanation seems to work because the neurally implemented mechanism *fits* with the relations that the contrastive questions let us discern among these closely related feelings.

Some may worry that such a structural explanation of the *relations* among entities or phenomena must of necessity fall short of explaining the constitution of these entities or phenomena themselves—which is what we set out to do (see, e.g., Levine 2001: 96, Chalmers 1996: 235). We respond to this by drawing on an analogy with other contrastive questions because, in general, contrastive explanations can in fact yield understanding of the constitution of the entities or phenomena that the contrasts concern. For example, we ask why water expands when freezing, rather than contracting as is the case with other liquids. The explanation cites facts about the constitution of water that differs from the constitution of the other liquids (when the kinetic energy drops sufficiently, the hydrogen bonds of H_2O help form an open lattice with a density less than that of liquid water, and other liquids do not have these hydrogen bonds). Here it does not make sense to

object that we have only explained something about relations between entities and nothing about constitution—it is obviously an explanation of something about the constitution of the process of freezing in water. If contrastive explanations in general can yield understanding of the constitution of things, then there is no reason to think that contrastive explanations of conscious states are barred from yielding understanding of the constitution of conscious states.

6. The philosophical challenge

The philosophical worry about whether neuroscience can explain consciousness is sometimes put in terms of the *intuition* that neuroscientific explanation seems to explain mere *belief* about conscious states, rather than the conscious state itself. For example, the story in terms of forward modelling and attenuation could seem to explain the belief that one is in control, but not how it *feels* to the subject—the famous 'what it is like' is missing. However, given our contrastive approach, we do not aim at explaining why there is a feeling of being in control rather than no feeling *at all*, only why it feels one way rather than another; and contrastive explanations should not be blamed for not explaining contrasts that they do not try to explain. Moreover, a major problem with this worry is that it rests on presuppositions that integrate very poorly with the influential and successful project of identifying the neural correlates of consciousness. The explanation about forward modelling and attenuation has its point of departure in the NCC of the feeling of being in control, not the NCC of the belief in being in control. It is difficult to see why, as the objection would have it, an account that rests on the NCC of *conscious* states suddenly becomes an explanation of mere *belief*. For this objection to be coherent, then, it would have to presuppose that the whole project of identifying the NCC is really only the project of identifying the neural correlates of belief states. But this is clearly a radical presupposition that integrates poorly with what we believe about the current scientific pursuit of the NCC. So we have in effect two candidate explanations: (a) the forward modelling and

attenuation mechanism rests on the NCC and explains aspects of the feeling of being in control, and (b) the mechanism rests on what we may call the neural correlates of belief and explains belief about being in control. Given that, as we have mentioned before, the best explanations are those that integrate well with what else we believe, we thereby have reason to believe that (a) is best and that it is indeed directed at conscious states.

However, some of the philosophical worries about the explanatory capacity of neuroscience stems from philosophical *arguments* concerning conceivability (e.g., Chalmers 1996). Applied to our case, the argument is that forward modelling and attenuation do not explain the conscious feeling of being in control because it is conceivable that some system, such as a simple robot, implements these mechanisms but without having the conscious feeling of being in control. If this is conceivable, then the forward modelling and attenuation story leaves it an open question whether that system is conscious. And an explanation is not successful if it leaves it an open question whether the thing to be explained occurred or not. The truth or falsity of the premises of this type of argument are hotly contested (see, e.g., Shear (ed.) 1998), but luckily the argument doesn't apply to the present suggestion because it ignores the contrastive character of the explanation. The right approach is, for example, to ask if the explanation leaves it an open question whether the system is feeling in control, rather than feeling merely that it predicted something; and we have attempted to show that *this* question isn't left open.

It might be objected that it is easily conceivable that a simple robotic arm can implement just those relevant mechanisms of forward modelling and attenuation without feeling anything whatsoever, and that therefore the explanation simply fails to address questions concerning consciousness—including contrastive questions—and thus couldn't even begin to close them off. The problem with

this objection is that it overgenerates and thus rules out all sorts of uncontroversial contrastive explanations. For example, it is plausible to assume that historians can explain why the great depression occurred in 1929 rather than in 1927 by alluding to salient events occurring in the socio-economic history between those two years. The above objection has the implausible consequence that such an explanation would be a failure because just these salient events *on their own* do not constitute a coherent stretch of socio-economic history at all and thus leaves it an open question whether the depression occurred in 1929, rather than 1927. The problem with the objection is that it ignores that, when trying to explain a particular contrast, we should be allowed to hold the system's other aspects constant, irrespective of how much we know or can explain about those other aspects.

Finally, it might be objected (1) that the philosopher's conceivability argument in its strongest form concerns *complete* physical and functional duplicates of us, and (2) that it is conceivable and thus possible that complete duplicates have no conscious states, and that therefore (3) it is left an open question whether the duplicates have conscious states. In response, we stress that our aim here, and the aim of most neuroscientists in general, is not to demonstrate the truth of materialism in one step and that we therefore are committed to no claim about complete duplicates. More likely, the truth of materialism will be established in many small steps, such as the one concerning the feeling of being in control, that each serve as inductive evidence for the materialist hypothesis (for some discussion, see Gold and Stoljar 1999)

We conclude that the philosopher's conceivability argument cannot show that the proposed account fails to be a constitutive explanation of aspects of a conscious state: the argument either misidentifies the target, overgenerates, or overcommits.

7. Concluding remarks

Cognitive neuroscience still has a very long way to go in the study of consciousness. But, in spite of some of the philosophical worries, it is false to say that neuroscience currently is wholly unexplanatory. Taking the studies of the neural correlates of conscious states as the point of departure, neuroscience can give contrastive explanations of conscious states. Once such explanations are at hand, they can—if good enough—help consolidate a given interpretation of the introspective reports used in the study of consciousness.

References:

Baars, B. 1997. Contrastive phenomenology: a thoroughly empirical approach to consciousness. In N. Block, O. Flanagan, G. Güzeldere (eds.), *The Nature of Consciousness*, Cambr. Mass.: MIT Press.

Baars, B. 2003. How brain proves mind: neuroimaging confirms the fundamental role of conscious experience. *Journal of Consciousness Studies* 10(9-10): 100-114.

Bird, A. 1998. *Philosophy of Science*. London: UCL Press.

Blakemore, S.-J., Wolpert, D.M., & Frith, C.D. 1998. Central cancellation of self-produced tickle sensation. *Nature Neuroscience* 1(7): 635-640

Blakemore, S.J., Frith, C.D., Wolpert, D.W. 1999a. Spatiotemporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience* 11: 551-559.

Blakemore, S.J., Wolpert, D., Frith, C.D. 1999b. The cerebellum contributes to somatosensory cortical activity during self-produced tactile stimulation. *NeuroImage* 10: 448-459.

Blakemore, S-J, Frith, C.D., Wolpert, D.W. 2001. The cerebellum is involved in predicting the sensory consequences of action. *NeuroReport* 12(9): 1879-85.

Blakemore, S.-J., Wolpert, D. M., Frith, C. 2002. Abnormalities in the awareness of action. *Trends in Cognitive Sciences* 6:237-242.

Blakemore, S-J, Oakley, DA & Frith, CD. 2003. Delusions of alien control in the normal brain. *Neuropsychologia* 41: 1058–1067.

Block, N. 1995. On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18: 227-287.

Block, N. 2001. Paradox and cross purposes in recent work on consciousness. *Cognition* 79:197-219.

Block, N., Stalnaker, R. 1999. Conceptual analysis, dualism and the explanatory gap. *Philosophical Review* 108:1-46.

Castiello, U., Paulignan, Y. & Jeannerod, M. 1991 Temporal dissociation of motor responses and subjective awareness. A study in normal subjects. *Brain* 114: 2639-2655.

Chalmers, D. 1996. *The Conscious Mind*. Harvard: Oxford University Press.

Chalmers, D. 2000. What is a neural correlate of consciousness? In *Neural Correlates of Consciousness: Empirical and Conceptual Issues*, T. Metzinger (ed.), Cambr. Mass: MIT Press.

Clark, A. 2000. Twisted tales: causal complexity and cognitive scientific explanation. In F. C. Keil & R. A. Wilson (eds.) *Explanation and Cognition*. Cambridge, Mass.: MIT Press. Pp. 145-166.

Cole, J. 1991. *Pride and a Daily Marathon*. London: Duckworth & Co.

Craver, C. F., Darden, L. 2001. Discovering Mechanisms in Neurobiology. In Machamer, P., Grush, R., McLaughlin, P. (eds.). 2001. *Theory and Method in the Neurosciences*. Pittsburgh: University of Pittsburgh Press.

Cummins, R. 2000. "How does it work?" versus "What are the laws?": Two conceptions of psychological explanation. In F. C. Keil & R. A. Wilson (eds.) *Explanation and Cognition*. Cambridge, Mass.: MIT Press. Pp. 117-144.

Decety J., Chaminade T., Grezes J., Meltzoff A.N. 2002. A PET exploration of the neural mechanisms involved in reciprocal imitation. *Neuroimage*, 15(1), 265-72.

Desmurget, M. & Grafton, S. 2000 Forward modelling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences* 4, 423-431.

Eliasmith, C. 2003. Moving beyond metaphors: Understanding the mind for what it is. *Journal of Philosophy* C: 493-520

Farrer, C. & Frith, C.D. 2002. Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency. *NeuroImage* 15(3): 596-603
Fletcher, P., McKenna, P.J., Friston, K. J., Frith, C.D., Dolan, R.J. 1999. Abnormal cingulate modulation of fronto-temporal connectivity in schizophrenia. *NeuroImage* 9: 337-342.
Fourneret, P. and Jeannerod, M. (1998) Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia* 36,
1133–1140.
Fourneret, P., Paillard, J., Lamarre, Y., Cole, J., Jeannerod, M. 2002. Lack of conscious recognition of one's own actions in a haptically deafferented patient. *NeuroReport* 13: 541-537.
Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Daléry, J., d'Amato, T. & Jeannerod, M. 2001. Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiat*ry 158: 454-459.
Friston, K. 2002. Functional integration and inference in the brain. *Progress in Neurobiology* 68: 113–143.
Frith, C. D. 1992. *The Cognitive Neuropsychology of Schizophrenia*. Hillsdale, NJ: Lawrence Erlbaum Ass.
Frith, C. D. 1995. Consciousness is for other people. *Behavioral and Brain Sciences* 18:4: 682-683.
Frith, C. D. 2002a. How can we share experiences? *Trends in Cognitive Sciences* 6:9: 374.
Frith, C. D. 2002b. Attention to action and awareness of other minds. *Consciousness and Cognition* 11: 481-487.
Frith, C.D. and Done, D.J. 1989 Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine* 19: 359-363.
Frith, C., Perry, R., Lumer, E. 1999. The neural correlates of conscious experience: an experimental framework. *Trends in Cognitive Sciences* 3:105-114.
Frith, C., Blakemore, S.-J., Wolpert, D. M. 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.-J., Wolpert, D. M. 2000b. Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews* 31: 357-363.
Gold, I. and Stoljar, D. 1999. A neuron doctrine in the philosophy of neuroscience. *Behavioral and Brain Sciences* 22: 809-69.
Goldberg, G., Mayer, N. H., Toglia, J. U. 1981. Medial frontal cortex and the alien hand. *Archives of Neurology* 38: 683-686.
Grush, R. forthcoming**.** The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*.
Haggard, P., Eimer, M. 1999. On the relation between brain potentials and awareness of voluntary movements. *Experimental Brain Research* 126: 128-133.
Haggard, P., Magno, E. 1999. Localising awareness of action with transcranial magnetic stimulation. *Experimental Brain Research* 127: 102-107.
Haggard, P. 2001. The psychology of action. *British Journal of Psychology* 92: 113–129.
Haggard, P., Clark, S., Kalogeras, J. 2002. Voluntary action and conscious awareness. *Nature Neuroscience* 5(4): 382-385.
Haggard, P., Martin, F., Taylor-Clarke, M., Jeannerod, M. and Franck, N. 2003. Awareness of action in schizophrenia. *Neuroreport* 14: 1081–1085.
Haggard P., Johnson H. 2003. Experiences of Voluntary Action. *Journal of Consciousness Studies* 10(9-10): 72-84.

Hempel, C. G. 1965. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.
Hohwy, J. and Frith, C. 2004. The Neural Correlates of Consciousness: Room for improvement, but on the right track: Response to Noë and Thompson. *Journal of Consciousness Studies* 11(1): 45-51.
Hohwy, J. 2004. Evidence, explanation, and experience: On the harder problem of philosophy. *The Journal of Philosophy* CI: 201-213.
Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Pütz, B., Yoshioka, T., Kawato, M. 2000. Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature* 403: 192–95.
Jack, A. I., Roepstorff, A. 2002. Introspection and cognitive brain mapping: from stimulus-response to script-report. *Trends in Cognitive Sciences* 6:8:333-339.
Jack, A. I., Shallice, T. 2001. Introspective physicalism as an approach to the science of consciousness. *Cognition* 79:161-196.
Jackson, F. 1986. What Mary didn't know. *Journal of Philosophy* 83:291-295.
James, W. 1890. *The principles of psychology.* New York: Holt
Kanwisher, N. 2001. Neural events and perceptual awareness. *Cognition* 79: 89-113.
Laureys, S., Faymonville, M. E., Luxen, A., Lamy, M., Franck, G.& Maquet, P. 2000 Restoration of thalamocortical connectivity after recovery from persistent vegetative state. *Lancet* 355:1790-1791.
Levine, J. 1983. Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly* 64:354-361.
Levine, J. 2001. *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.
Lewis, D. 1986. Causal explanation. In *Collected Papers, Vol. II*. Oxford: Oxford University Press. Pp. 214-240.
Libet, B., Gleason, C. A., Wright, E.W., Pearl, D.K. 1983. Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): the unconscious initiation of a freely voluntary act. *Brain* 106: 623-642.
Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences* 8: 529-566.
Lipton, P. 1991. *Inference to the Best Explanation*. London: Routledge.
Lipton, P. 1993. 'Contrastive Explanation'. In David-Hillel Ruben (ed.), *Explanation*, Oxford: Oxford University Press.
Machamer, P., Grush, R., McLaughlin, P. (eds.). 2001. *Theory and Method in the Neurosciences*. Pittsburgh: University of Pittsburgh Press.
Machamer, P., Darden, L., Craver, C. F. 2000. Thinking about mechanisms. *Philosophy of Science* 67: 1-25.
McCloskey, D.I., Colebatch, J.G., Potter, E.K., Burke, D. 1983. Judgements about the onset of rapid voluntary movements in man. *Journal of Neurophysiology* 49: 851-863.
Mellors, C. S. 1970. First-rank symptoms of schizophrenia. *British Journal of Psychiatry* 117: 15-23.
Meltzoff, A.N. & Decety, J. (2003) What imitation tells us about social cognition: a rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 358: 491-500.
Mill, J. S. 1904. *A System of Logic*. London: Longmans, Green & Co.
Newton-Smith, W. 2000. Explanation. In W. Newton-Smith (ed.) *A Companion to Philosophy of Science*. Oxford: Blackwell.

Noë, A., Thompson, E. 2004. Are there neural correlates of consciousness? *Journal of Consciousness Studies* 11(1): 3-28.
Portas, C., Maquet, P., Rees, R., Blakemore, S-J. & Frith, C.D. 2003. The neural correlates of consciousness. Chapter 16 in R.S.J. Frackowiak et al. (eds) *Human Brain Function. 2nd Edition.* San Diego, CA: Elsevier.
Russell, J. 1996. *Agency: Its Role in Mental Development.* Hove East Sussex: Erlbaum (UK) Taylor Francis.
Schooler, J. 2002. Re-representing consciousness: dissociations between experience and meta-experience. *Trends in Cognitive Sciences* 6:8: 339-344.
Shear, J. (ed.). 1998. *Explaining consciousness : the "hard problem"*. Cambridge, Mass. : MIT Press.
Sirigu A, Daprati E, Pradat-Diehl P, Franck N & Jeannerod M. 1999. Perception of self-generated movement following left parietal lesion. *Brain* 122: 1867-74.
Spence, S.A., Brooks, D.J., Hirsch, S.R., Liddle, P.F., Meehan, J. & Grasby, P.M. 1997. A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain* 120: 1997-2011.
Stenneken, P., Aschersleben, G., Cole, J., Prinz, W. 2002. Self-induced versus reactive triggering of synchronous movements in a deafferented patient and control subjects. *Psychological Research* 66: 40-49.
Weiller C., Juptner M., Fellows S., Rijntjes M., Leonhardt G., Kiebel S., Muller S., Diener H.C., Thilmann A.F. 1996. Brain representation of active and passive movements. *NeuroImage* 4(2): 105-10.
Wegner, D. and Wheatley, T. 1999. Apparent mental causation: sources of the experience of will. *American Psychologist* 54: 480-492.
Wolpert, D. M., Miall, R. C., Kawato, M. 1998. Internal models in the cerebellum. *Trends in Cognitive Science* 2(9): 338-347.
Wolpert., D. M., Ghahramani, Z. 2000. Computational principles of movement neuroscience. *Nature Neuroscience* Supplement 3: 1212-1217.
Wolpert, D., Ghahramani, Z., Jordan, M. I. 1995. An internal model for sensorimotor integration. *Science*: 269: 1880-1882.
Wolpert, D. M., Doya, K., Kawato, M. 2003. A unifying computational framework for motor control and social interaction.*Philosophical Transactions of the Royal Society London B* 358: 593–602.