

Phenomenology and the Feeling of Doing: Wegner on the Conscious Will¹

Tim Bayne, Macquarie University

tbayne@scmp.mq.edu.au

February 24, 2005

To appear in S. Pockett, W. P. Banks & S. Gallagher (eds) *Does Consciousness Cause Behavior? An Investigation of the Nature of Volition*. Cambridge, MA: MIT Press.

Abstract

In this paper I examine Daniel Wegner's claim that the conscious will is an illusion. I begin by asking what Wegner might mean by "the conscious will", and what he might mean by calling it an illusion. I then explore Wegner's arguments for the claim that the conscious will is an illusion. I conclude that these arguments fail, and that although there is much we do not yet know about the structure of the phenomenology of agency, it is unlikely that the phenomenology of agency is illusory.

1. Introduction

Given its ubiquitous presence in everyday experience, it is surprising that the phenomenology of doing—the experience of being an agent—has received such scant attention in the consciousness literature. But things are starting to change, and a small but growing literature on the content and causes of the phenomenology of first-person agency is beginning to emerge.² One of the most influential and stimulating figures in this literature is Daniel Wegner. In a series of papers and his book *The Illusion of Conscious Will* (ICW) Wegner has developed an account of what he calls "the experience of conscious will". In this paper I assess Wegner's model of the conscious will and the claims that he makes on the basis of it. Although my focus is on Wegner's work, many of the issues I raise are relevant to the study of the phenomenology of agency more generally.

¹ An earlier version of this paper was presented at a workshop on the phenomenology of agency organized by Terry Horgan at the University of Arizona and a workshop at Johannes Gutenberg-Universität, organized by Thomas Metzinger. I am grateful to the audiences on both occasions for their comments. I am also grateful to David Chalmers, Tim Crane, Frederique de Vignemont, Patrick Haggard, Neil Levy, Al Mele, Thomas Metzinger, Eddy Nahmias, Elisabeth Pacherie, and Sue Pockett for very helpful comments on this paper. This work was supported by Australian Research Council Discovery Grant DP0452631.

² Recent examples of this work include Haggard et al (2004), Haggard (in press), Haggard & Johnson (2003); Hohwy (in press); Metzinger (in press), Nichols (2004), Horgan, Tienson and Graham (2003), Nahmias, Morris, Nadelhoffer & Turner (2004), Pockett (2004).

I regard Wegner's work on the phenomenology of agency as comprised of two components. First, Wegner provides a model of how the experience of conscious will is generated. Roughly speaking, Wegner's view is that we experience conscious will with respect to an action when we have an introspective preview of it.

The experience of consciously willing our actions seems to arise primarily when we believe our thoughts have caused our actions. This happens when we have thoughts that occur just before the actions, when these thoughts are consistent with the actions, and when other potential causes of the actions are not present. (Wegner 2005: 23)

I will call this the *matching model* of the experience of conscious will.

The second component of Wegner's account is his claim that the conscious will is an illusion. Exactly what Wegner might mean by describing the conscious will as an illusion is open to some debate (see § 4), but I take his central claim to be this: the experience of conscious will misrepresents the causal path by means of which one's own actions are generated.

How are the two components of Wegner's work related? The most straightforward way to read Wegner is to take him to be offering the matching model as evidence for the claim that the conscious will is an illusion. I will argue that if this is indeed Wegner's view then it is mistaken. The matching model might explain why people experience a sense of agency, but it does not show that this sense of agency is an illusion. I proceed as follows. In the following section I explore the notion of the conscious will. In section 3 I examine Wegner's matching model of the conscious will, and in sections 4 and 5 I turn my attention to his claim that the conscious will is an illusion.

2. What is the conscious will?

The notion of the experience of conscious will is anything but straightforward, and there is ample room for terminological confusion in these waters. One issue concerns the relationship between the experience of conscious will and the will itself. It is natural to suppose that acts of the will are the intentional objects of experience of conscious will. On this view, the logical relationship between experiencing oneself as willing something and actually willing it is akin to the relationship between, say, experiencing something as being blue and it actually being blue. The experience is one thing, the property or object experienced is another. Although much of what Wegner says can be read in these terms, there are passages in his work which suggest a very different conception of the relationship between experience of the will and the will itself. For example, Wegner states that "will is a feeling, not unlike happiness or sadness or anger or anxiety or disgust" (ICW, 326; see also ICW, 3 and ICW, 29). But referring to the will as a feeling threatens to collapse the distinction between the will itself and experiences of the will.

A second issue concerns the representational content of the experience of conscious will. For the most part Wegner leaves the notion of the experience of conscious will at an intuitive level, often describing it simply as 'the experience of doing'. He does, however, link the experience of conscious will with various other aspects of the phenomenology of agency, such as the experience of *authorship* (Wegner 2005, 27); the experience of

intentionality (Wegner and Erskine 2003: 688); the experience of *effort* (ICW, 39); the experience of experience of *free will* (ICW, 318); and the experience of *mental causation* (Wegner 2005, 23).

This terminological proliferation presents us with something of a challenge. Are these terms meant to be (rough) synonyms for a single type of experience, or are they meant to refer to distinct experiential types? Wegner does not tell us. One could of course stipulate that these terms are being used synonymously, but Wegner does not appear to have made this stipulation, and so we are left to rely on our pre-theoretical grasp of these notions. Although the recent writers on the phenomenology of agency seem to assume that at least some of these terms are synonyms it is far from obvious that this is the case.³ *Prima facie*, at least some of these terms appear to refer to different experiential contents. To take just one contrast, the experience of authorship appears to differ from the experience of effort (see Bayne and Levy, in press). But if these various terms refer to distinct types of experiences then it is entirely possible that the matching model might apply to some of them but not others. Furthermore, it might turn out that some of these experiential types are illusory but others are not.

Wegner appears to regard the experience of conscious will as most intimately associated with two kinds of experiences: the experience of “mental causation” and the experience of authorship. Let me take these two notions in turn.

Wegner often refers to his account of the conscious will as “the theory of apparent mental causation”. He writes: “Intentions that occur just prior to action ... do seem to compel the action. This is the basic phenomenon of the experience of will” (ICW, 20). The view, I take it, is that an experience (as) of conscious will *just is* an experience (as) of one’s intentions causing the action in question.⁴ Chris Frith endorses a similar view: “Our sense of being in control derives from the experience that our actions are caused by our intentions” (Frith, 2002: 483).

But is there really such a thing as the experience of mental causation? To put the question slightly differently, does the *normal* phenomenology of agency include an experience of mental causation? The experience of being an agent is certainly imbued with a sense of intentionality and purposiveness—we experience ourselves acting in light of our beliefs and desires and as realizing our intentions—but is this experience really *causal*? A number of authors have suggested that it is not (see Horgan, Tienson, & Graham, 2003; Searle 2001; Wakefield & Dreyfus 1991). We do not typically experience ourselves as being *caused* to move by our intentions. Arguably, if there is any causal component to the phenomenology of agency (and it is not clear to me that there is) it involves an

³ Consider the following passage: “Most of us navigate through our daily lives with the belief that we have free will: that is, we have conscious intentions to perform specific acts, and those intentions can drive our bodily actions, thus producing a desired change in the external world” (Haggard and Libet, 2001: 47). I think this equation of the free will with mental causation is implausible. It leads to the counter-intuitive conclusion that those who deny the reality of free will are thereby forced to deny the reality of mental causation.

⁴ This is the view suggested by *most* of Wegner’s work. However, in a recent paper Wegner suggests that intention confabulation can occur when we ascribe intentions to ourselves in order to account for our experience of willing an action (Preston and Wegner 2005). This suggests that Wegner may now think that we can have an experience of conscious will in the *absence* of an experience of mental causation.

experience of oneself, rather than one's mental states, as the cause of one's movements.⁵ At any rate, the experience of will does not seem to involve an experience of one's intentions *compelling* one's actions. Leaving aside questions about whether the experience of "mental causation" is really causal, it is certainly true that it occupies a central place in the phenomenology of agency.

Consider now to the experience of authorship. I'm not sure whether Wegner thinks of the experience of authorship as identical to (or included within) the experience of conscious will, but he does suggest that the matching model of the experience of conscious will also functions as a model of the experience of authorship: "When the right timing, content, and context link our thought and our action, this construction yields a feeling of authorship of the action. It seems that we did it" (Wegner 2005: 27).

What exactly *is* the feeling of authorship? The feeling of authorship, I take it, is the experience that you yourself are the agent or author of an event – the doer of the deed. The contrary of the feeling of authorship is the feeling that the event was done by someone else, or had no agent as its source. But although it is helpful as an intuitive starting point, this gloss on the notion of authorship only gets us so far; we also need to know what the satisfaction conditions of experience of authorship are. What is it to author an action?

The quandaries that confront us here can be illustrated by considering Penfield actions (Penfield 1975). When Wilder Penfield directly stimulated the exposed brains of conscious patients he triggered a variety of actions. But Penfield's patients reported that they did not "do" the action in question, and instead felt that Penfield had pulled it out of them. Wegner advances Penfield actions as showing that it is possible to author an action without experiencing oneself as authoring it (ICW: 47). But should we accept that Penfield actions were authored by Penfield's patients? Perhaps they were authored by Penfield. Or perhaps they were not authored at all. More generally, there is a range of pathologies—utilization behaviour, the anarchic hand syndrome, and hypnotically-induced movement—where there is some obscurity about whether the subject in question stands in the authoring relation to the movement. But if we are in the dark as to when an action is authored then we are also in the dark as to when experiences of authorship are veridical, and to the extent that the experience of conscious will includes the experience of authorship, we are also in the dark as to when the experience of conscious will is veridical.

The question of what Wegner might mean by the experience of conscious will is further complicated by the distinction between willed and unwilled actions (see e.g. Jahanshahi and Frith 1998). Wegner himself invokes this distinction, often describing willed actions as "controlled" (or "voluntary") and unwilled actions as "automatic". The distinction between willed and unwilled actions is roughly the distinction between those actions that derive from the agent's plans and goals and those actions that are stimulus-driven. (I say "roughly", for at times the distinction between controlled and automatic actions seems to involve the rather different distinction between those actions that are accompanied by a sense of doing and those that are not.) Given the distinction between willed and automatic action one might wonder whether Wegner is using "the experience of

⁵ See Horgan et al (2003) and Nahmias et al (2004) for some useful discussion of these issues.

conscious will” to refer to an experience of the target action as willed rather than automatic. But I think that this cannot be what Wegner has in mind. On this reading, the claim that the experience of conscious will is an illusion would amount to the claim that there are no willed (plan-driven) actions - all actions are automatic. Not only is this an implausible reading on its own terms, it also runs counter to Wegner’s own work on the phenomenon of mental control. Just how the experience of conscious will might be related to the distinction between controlled and automatic actions is, I think, something of an open question.

We have seen that Wegner draws a close connection between the experience of conscious will and the experiences of mental causation and authorship. This gives us an intuitive grip on what the experience of conscious will might be, but it also raises unresolved questions. Not only are there questions about exactly how the experiences of mental and mental “causation” are related to the experience of conscious will (and to each other), there are also questions about how the experiences of mental causation and authorship should themselves be understood. Much more could be said here, but I turn now to Wegner’s account of how the experience of conscious will might be produced.

3. The matching model

According to Wegner’s matching model,

The experience of consciously willing our actions seems to arise primarily when we believe our thoughts have caused our actions. This happens when we have thoughts that occur just before the actions, when these thoughts are consistent with the actions, and when other potential causes of the actions are not present. (Wegner 2005: 23)

There are three questions that can be raised here. (1) What kind of thoughts are involved in generating the experience of will? (2) What kind of consistency is needed between the target thoughts and the target actions? (3) How is the presence of “other potential causes of the action” incorporated into the model? Let me take these questions in turn.

Matching models can be pitched at a number of explanatory levels. Some are *sub-personal*; that is, the target thoughts are not fully-fledged mental states, such as beliefs or intentions, but sub-personal states, such as motor commands (see e.g. Blakemore & Frith, 2003; Haggard & Johnson 2003). Other matching models are pitched at a *personal* level, where it is a match between the agent’s beliefs about what they are doing (or their intentions to do such-and-such) and their awareness of what they are doing which generates the experience of agency. Wegner’s model operates at a personal-level of explanation. He says that the experience of conscious will is generated when the prior thought appears in consciousness, which suggests that the conscious will is generated by full-blown mental states of the person—intentions, presumably—rather than sub-personal states of the motor system (see e.g. Wegner 2003b, p.67).

But what exactly is the relationship between the awareness of a match between one’s intentions and one’s actions (on the one hand) and the experience of doing? Consider the following thought experiment (inspired by Davidson 1973):

Chloe wants to rid herself of the weight and danger of holding another climber on a rope, and she knows that by loosening her hold on the rope she could satisfy this desire. This belief and want leads her to form the intention to loosen her hold on the rope, but before she has the chance to act on this intention the contemplation of it so unnerves her that she loses her grip on the rope and the climber she is holding falls to his death. Chloe might experience the loosening of her hold as caused by her intention to loosen her hold, but she does not experience the loosening of her hold as an action – she does not experience this event as something that she did.

The case of Chloe the climber appears to show that one can experience one's movements as caused by (and realizing) one's intentions without experiencing a sense of agency towards them. It seems that the experience of doing cannot simply be generated by an awareness of a match between one's intentions and one's movements. Something more is needed. What might that be?

There are a number of ways in which one might respond to the case of Chloe. I have space here to examine just one line of response. Perhaps Chloe fails to experience a sense of agency with respect to her movements because she is not aware of them as implementing *fine-grained* intentions. Chloe experiences herself as having an intention to let go of the rope, but she is not aware of herself as having an intention to let go of the rope *in any particular way*. Note that this response incurs a certain cost. It suggests a model of the feeling of doing on which this feeling depends on awareness of *movement-directed* intentions, rather than an awareness of *goal-directed* intentions. And it is doubtful whether we are normally conscious of having movement-direct intentions. We are normally aware of the goals of our actions, but we are not normally aware of how we are intending to move our body in order to realize those goals.⁶ So the proponent of Wegner's matching model might want to explore other responses to the case of Chloe the climber.

What light do Wegner's experiments shed on the nature of target thoughts?⁷ Consider Wegner and Wheatley's I-Spy experiment (Wegner & Wheatley, 1999). In this experiment participants and an experimental confederate had joint control of a computer mouse that could be moved over any one of a number of pictures on a board (e.g., a swan). When participants had been primed with the name of the picture on which the

⁶ Note that by focusing on movement-direct intentions rather than object-direct intentions Wegner might be able to respond to an objection to his model made by Zhu (2004). Zhu points out that one can experience a sense of agency with respect to failed actions. There is something it is like to try (but fail) to push over a heavy boulder. But if one can have a sense of agency with respect to failed intentions, then the sense of agency cannot rest on an awareness of a match between one's intentions and one's actions. In response to this objection one might point out that even in failed attempts at action one is usually aware that one has successfully implemented an intention to move one's muscles, even if one has failed to achieve the goal state (moving the boulder). Of course, this is only a first step in answering Zhu's objection. Perhaps it is possible to experience a sense of agency without having any sense of successfully implementing an intention. I suspect that some episodes of mental agency might have this form.

⁷ See Hohwy (in press), Malle (this volume) and Pacherie (this volume) for further discussion of the I-Spy experiment.

mouse landed they showed a slightly increased tendency to self-attribute the action, even when the landing was guided by the confederate and not by the participant. Did the subjects in this experiment experience themselves as having the intention to land on the primed object, as some of Wegner's commentators have suggested (Blakemore and Frith, 2003: 220)? I see no reason to think that this was the case.⁸ Rather than assume that participants formed an intention to land on the swan it seems to me that one need only suppose that their representation of swans was activated.

Wegner's automatic typing priming experiment (reported in Wegner 2003a: 4) also points towards an account of the experience of conscious will that need not involve intentions. Participants in this experiment were primed with various words (such as "deer"), and then required to type letters randomly without seeing the computer screen. They were then asked to rate a series of words, none of which they had actually written, according to the degree to which they felt they had authored them. Wegner and colleagues found that participants were more likely to judge that they had authored the presented word when it had been a prime than when it had not been presented as a prime. Again, this experiment suggests that one can experience a sense of agency (at least retrospectively) without experiencing oneself as having (had) an intention to perform the action. Presumably the subjects experienced a sense of agency because they had had thoughts with content that had some semantic connection with the action.⁹

But there is, I suspect, a good reason why Wegner's presents his model in terms of matches between intentions and actions. We have a clear notion of what it is for an intention to match an action: an intention matches an action when and only when the action realizes (satisfies) the intention. It is less clear what it is for "thoughts" which consist merely of heightened activation of certain concepts to match an action. Perhaps the natural thing to say here is that a semantic activation matches an action when they have common content. But this, of course, won't do as an account of the phenomenology of agency. Consider someone who experiences himself as intending not to have a hamburger for lunch, only to find his hands reaching for a juicy hamburger. Despite common semantic content between the intention and the target action, the person's actions are unlikely to be accompanied by the feeling of doing. In short, by moving away from a conception of target thoughts as intentions we problematize the matching component of the matching model.

Finally, let me turn to the 'exclusivity condition' – the claim that the experience of conscious will is heightened when other potential causes of the actions are not present. Wegner includes the exclusivity condition in order to account of the finding that experiences of authorship can be modulated by the awareness of rival causes of agency (see Wegner 2003a, p. 5f.). But there is also something puzzling about including the exclusivity condition as part of the basic model of how the experience of consciousness will is generated, for in order to apply the exclusion condition an agent must already know whether or not someone (or something) else is a potential cause of the action in

⁸ Furthermore, even if participants had an intention to land the mouse on the swan, there is no reason to think that this intention was *conscious*.

⁹ This experiment also raises questions about what the other relatum of Wegner's matching relation is, for participants in this experiment did not have a perceptual experience of the action in question, they were only presented with an image of the results of the action.

question. Consider a schizophrenic suffering from delusions of control, who reports that God (or the Government, or the Devil) is causing him to raise his hand. Why does this person regard these (unobservable) agents as potential causes of his movements, and why does he regard them as exercising this potentiality on this particular occasion? This, it seems to me, is the really difficult question for accounts of the phenomenology of agency. Normally, of course, we do not regard God (or the Government, or the Devil) as potential causes of our movements. (And if, like Malebranche, we do regard these agents as potential causes of our movements, this does not undermine our experience of authorship.) All the heavy lifting appears to be done by an account of those systems responsible for identifying an entity as a potential cause of one's movements, and the principles by means of which we regard one agent as a more plausible cause of our actions than another. Of course, it is no criticism of Wegner that he fails to provide such an account, but insofar as these questions go unanswered any model of authorship-processing will be incomplete.

4. What is illusory about the experience of conscious will?

I turn now to Wegner's claim that the conscious will is an illusion. Wegner's work suggests more than one reading of this claim.¹⁰ Consider the following passage:

We often consciously will our own actions. This experience is so profound that it tempts us to believe that our actions are caused by consciousness. It could also be a trick, however – the mind's way of estimating its own apparent authorship by drawing causal inferences about relationships between thoughts and actions. Cognitive, social and neuropsychological studies of apparent mental causation suggest that experiences of conscious will frequently depart from actual causal processes and so might not reflect direct perceptions of conscious thought causing action. (TICS, abstract)

Three ideas seem to be contained in this passage: (1) We experience our actions as caused by consciousness, but this experience is non-veridical (misleading); (2) We have a mistaken model of how experiences of conscious will are generated: we think that they are based on direct perception of their objects, but they are in fact theoretically-mediated; (3) Experiences of conscious will are frequently non-veridical. I will examine all three claims. The first two can be dealt with fairly quickly; the third will prove more troublesome.

4.1 Do we experience our actions as caused by consciousness?

Do we believe that our own actions are caused by consciousness? The first thing to ask here is: consciousness *of what*? Wegner appears to have consciousness of agency in mind. For example, he says that although we believe that our experiences of conscious will cause our actions, they do not (see, for instance, ICW, 318). Do we believe that experiences of conscious will cause our actions? More precisely, do we believe that the experience of consciously willing an action causes the movement (or action) that is the target of that experience?

¹⁰ For further discussion of what Wegner might mean by calling the conscious will an illusion see Nahmias (2002) and Holton (2004).

Some people appear to believe this. Searle says that it is part of the intentional content of an experience of acting that the experience itself causes the bodily movement in question (Searle 1983, p. 124; see also Searle 1983, p. 88, 95). I do not share Searle's view. It does not seem to me that my experience of raising my hand causes my hand to go up (or causes me to raise my hand). I might perhaps experience my hand as rising in virtue of my intention to raise it, but I do not experience my hand as rising in virtue of my *experience* of my intention to raise it. (Nor do I experience it as rising in virtue of my having an experience of agency towards its motion.)

Note that nothing in what I have just said involves a commitment to the view that consciousness is epiphenomenal. (Nor, of course, does it involve a rejection of epiphenomenalism!) To deny that experiences of agency bring about their target actions is not to deny that experiences of agency are causally efficacious – indeed, it is not even to deny that we experience experiences of agency as causally efficacious. It is perfectly possible that experiences of agency play various causal roles—such as generating beliefs about what we are doing. The point is that they do not cause those movements that are their intentional objects; or at least, if they do, we do not experience them as doing so.

4.2. The experience of the will as theoretically mediated

A second reason Wegner gives for rejecting experiences of conscious will as illusory is that they are theoretically mediated. As he puts it, “[Conscious will is an illusion] in the sense that the experience of consciously willing an action is not the direct indication that the conscious thought has caused the action.” (ICW, 2) It is not clear to me what “direct indication” might mean in this context. One possibility is that a direct indication view of the phenomenology of agency involves thinking that one can eyeball agency directly: that one's experiences of (and beliefs about) one's own agency involve no sub-personal processing or inferential mechanisms of any kind. I suppose that some folk might hold such a view of the phenomenology of (first-person) agency, but I do not know of any one who does, and I do not know of any reason for thinking that this view is plausible. More to the point, I do not know of any reason for thinking that experiences of agency could be reliable only if the direct indication theory were true. The rejection of direct indication views of visual perception has not led theorists to reject visual experiences as illusory, nor should it have. I don't see why the rejection of direct indication accounts of the phenomenology of agency should lead us to reject the phenomenology of agency as an illusion.

5. The conscious will as non-veridical

I turn now to the third and, I think, central sense in which Wegner thinks of the conscious will as an illusion. On this view, the conscious will is an illusion in the sense that experiences of doing are systematically, or at least frequently, *non-veridical*: Experiences of doing misrepresent our agency and the structure of our actions. I detect two lines of argument in Wegner's work for this thesis. The main argument involves dissociations between the phenomenology of agency and the exercise of agency. A subsidiary argument involves an appeal to eliminativism. I begin with the argument from dissociation.

5.1. The Dissociation Argument

The Illusion of Conscious Will contains extensive discussion of dissociations between the exercise of agency and the phenomenology of agency. Some of these cases appear to demonstrate that we can experience ourselves as doing something that someone else is doing (and that we are not). Wegner calls such cases *illusions of control*. The I-Spy experiment, discussed earlier, is an example of an illusion of control. Other dissociations involve experiencing someone (or something) else as the agent of what one is doing. Wegner calls such phenomena *illusions of action projection*. Among the most fascinating of the various illusions of action projection that he discusses is facilitated communication, a practice that was introduced as a technique for helping individuals with communication difficulties. Facilitators would rest their hands on the hands of their clients as the client typed a message. Although the facilitators experienced no sense of authorship towards the typed message, there is ample evidence that the content of “facilitated” messages derived from the facilitator rather than the client (see Felce 1994; Jacobson, Mulick and Schwartz 1995; Wegner, Fuller, and Sparrow, 2003).

Although these dissociations clearly play an important role in Wegner’s argument for the illusory nature of the conscious will, it is possible to provide different interpretations of that role. On one reading Wegner is mounting an inductive generalization: since some experiences of conscious will are non-veridical it is reasonable to infer that most, and perhaps even all, such experiences are.¹¹ The problem with this argument is obvious: the fact that experiences of agency *can* be non-veridical shows that the mechanisms responsible for generating such experiences are fallible, but it does not show that they are unreliable.

Another way to read the argument from dissociations is as an inference to the best explanation.¹² The argument proceeds as follows: Since the phenomenology of agency plays no direct role in the genesis of action where such experiences are absent, we have good reason to think that it plays no direct role in the genesis of action when such experiences are present. This idea, I take it, is the thought lying behind the following claim: “If conscious will is illusory, automatisms are somehow the ‘real thing’, fundamental mechanisms of mind that are left over once the illusion has been stripped away. Rather than conscious will being the rule and automatism the exception, the opposite may be true.” (ICW, 143; see also ICW, 97).

I share Wegner’s suspicion that automatisms may, in some sense, be ‘the real thing’ – the ‘fundamental mechanisms of mind’. But I don’t share Wegner’s view that this result would be inconsistent with the phenomenology of agency. Of course, much depends on what precisely it means to say that automatisms are the fundamental mechanisms of mind. To the extent that automatisms are action-generation procedures that do not involve intentional states of any kind then there may be a tension between automaticity and the experience of conscious will, but Wegner provides little evidence for the view that our actions are usually automatistic in this sense of the term. If, on the other hand, automatisms are action-generating procedures that are non-consciously initiated then

¹¹ Bayne (2001) presents Wegner’s argument in these terms. Nahmias (2002) also suggests that one can read Wegner as advancing an inductive generalization of this kind.

¹² I thank Sue Pockett for drawing this reading of Wegner’s work to my attention.

there is ample reason to describe much of what we do as automatic in nature. But on this conception of an automatism there is, I claim, no conflict between automaticity and the feeling of doing. So there is no argument from automaticity (thus conceived) to the claim that the experience of conscious will is an illusion.

I do not deny that the phenomenology of agency *can* be illusory. Consider, for example, the experience of intentionality. An experience of intentionality will be non-veridical if the action in question is not guided by an intention, or if it is guided by an intention other than the one that it seems to have been produced by. The phenomenon of confabulation suggests that at least one if not both of these conditions occur. But I think that there is little reason to assume that either kind of mistake is at all common in everyday life. Confabulation is striking precisely because it is unusual.

In fact, Wegner's own account of the genesis of the experience of doing suggests that such experiences will normally be veridical. According to the matching model, we experience ourselves as doing X when we are aware of our intention to X as being immediately prior to our X-ing, and when we are not aware of any rival causes of our X-ing. Now, if we experience ourselves as having an intention to X then it probably is the case that we do have the intention to X. (After all, it seems reasonable to suppose that introspection is generally reliable. At any rate, Wegner is in not position to challenge the reliability of introspection, for he himself assume the reliability of introspection insofar as he takes subjects to be reliable in reporting their experiences of conscious will.) But if one has an intention to X, and if one has in fact X-ed, and if one's intention to X is immediately prior to one's X-ing, then it is highly likely that one's intention to X is involved in bringing about one's X-ing. It would be remarkable if one had an intention to raise one's hand just prior to raising one's hand but the intention played no causal role in the raising of one's hand. Far from showing that experiences of agency are illusory, Wegner's own model of how such experiences are generated predicts that they will normally be veridical.

5.3. The argument from eliminativism

There is a further and quite different line of argument for thinking that experiences of the conscious will are non-veridical to be found in Wegner's work. I call it the argument from eliminativism. Consider the following passages:

... it seems to each of us that we have conscious will. It seems we have selves. It seems we have minds. It seems we are agents. It seems we cause what we do. *Although it is sobering and ultimately accurate to call all this an illusion*, it is a mistake to conclude that the illusion is trivial. On the contrary, the *illusions piled atop apparent mental causation* are the building blocks of human psychology and social life. (ICW, 342; emphasis added)

The agent self cannot be a real entity that causes actions, but only a *virtual* entity, an *apparent* mental causer. (Wegner 2005: 23; emphasis added)

The theory of apparent mental causation turns the everyday notion of intention on its head...The theory says that people perceive that they are intending and that they understand behavior as intended or unintended – but they do not really intend. (Wegner 2003a; 10)

There are two closely related thoughts here: (1) the experience of conscious will must be an illusion because there are no selves; (2) the experience of conscious will must be an illusion because there is no intentional (mental) causation.

How committed Wegner is to either of these claims is something of an open question, for he also says that the conscious will is a “fairly good sign” of our own causation (ICW, 333, see also ICW, 334 and ICW, 327). How, one wonders, could the experience of conscious will be a “fairly good sign” of our own causation if we are merely virtual entities? There seems to be a deep tension between Wegner’s eliminativism and his willingness to allow that experiences of conscious will are often veridical.

This tension can be resolved by regarding Wegner as ambivalent about how much to build into the conception of the self as it is represented in the experience of authorship. In some passages Wegner seems to be working with a fairly weak conception of authorship, on which the experience of authorship merely tracks those actions that are mine and those that are produced by other people. Those passages in which Wegner claims that the experience of conscious will is generally reliable might be understood along these lines. After all, we are pretty good at working out whether we are raising our arms or someone else is raising them for us, and it would be hard to take an argument for the denial of this claim seriously. In other passages, however, Wegner appears to be working with a much stronger conception of authorship, according to which an experience of authorship represents one’s actions as caused by a homunculus of some kind: a substantial entity inside one’s brain that is the ultimate (uncaused) cause of one’s actions. And here, of course, there is ample room for eliminativist worries. There are no homunculi; and we are not ultimate causes.

But in acting do we experience ourselves as homunculi – as uncaused causes? I think the answer to this question is quite unclear. There is certainly a tradition of thought within philosophy that would reconstruct the phenomenology of authorship in terms of agent causation – that is, in terms of the experience of oneself as a *sui generis* cause. According to Timothy O’Connor, “[agent causation] is appealing because it captures the way we experience our own activity. It does not seem to me (at least ordinarily) that I am caused to act by the reasons which favour doing so; it seems to be the case, rather, that I produce my own decisions in view of those reasons...” (1995, p. 196). And although he himself does not endorse agent causation, Carl Ginet also gives an agent causal gloss on the phenomenology of agency: “My impression at each moment is that *I* at that moment, and nothing prior to that moment, determine which of several open alternatives is the next sort of bodily exertion I voluntarily make” (1990, p. 90; emphasis in original).

One might attempt to characterize the sense in which we experience ourselves as causes in a manner that does not involve any commitment to agent causation – or, at least, in such a way that does not involve any commitment to the idea that the agent cause must be an homunculus. Perhaps we experience our actions as caused by, well, ourselves, where this experience makes no commitment to any particular view of what kinds of things we are. Rather than conceive of the agent causal relation as holding between a Cartesian self and its actions, perhaps the agent causal theorist should think of it as holding between an organism and its actions (see e.g. Bishop 1983). If such a view can be defended, perhaps it is also possible that agent causal accounts of the experience of authorship can resist the threat of eliminativism.

Of course, we still need an account of when agents stand in the authoring relation to their movements. Consider again the phenomenology of the anarchic hand, in which the patient experiences fails to experience any sense of authorship with respect to the movements of her anarchic hand. Is this patient the author of these movements? Do these movements constitute *actions*, and if so, are the actions *her* actions? Her movements are clearly caused by events within her, but are they caused in such a way as to make it the case that her movements constitute an action that she performs? These are very difficult questions. Not only is it unclear what the answers to them are, it is also unclear how we ought to go about answering them. But unless we have some handle on them we don't really have an adequate grip on the phenomenology of authorship.

Much more could be said about eliminativism and the experience of authorship, but I turn now to the topic of eliminativists about mental causation. Consider the following passage:

When we apply mental explanations to our own behavior-causation mechanisms, we fall prey to the impression that our conscious will causes our actions. The fact is, we find it enormously seductive to think of ourselves as having minds, and so we are drawn into an intuitive appreciation of our own conscious will.... The real causal sequence underlying human behavior involves a massively complicated set of mechanisms. (ICW, 26f.)

To contrast intentional causation with 'real causal sequences' is to suggest that intentional causation is not real – that it is merely ersatz or virtual. Wegner's argument for this view seems to be a version of the argument from explanatory exclusion: lower levels of explanation exclude (eliminate) higher levels of explanation.

There is a lively debate in the philosophy of mind about what, if anything, is wrong with the exclusion argument (see e.g. Block 2003; Kim 1998). The debate is a complex one, and here is not the place to enter into it. Suffice it to say that it proceeds against the backdrop of a general—although far from universal—commitment to the causal efficacy of mental content.¹³ Mental causation is no more undermined by the presence of underlying neuropsychological processes than neuropsychological causation is undermined by the presence of underlying neurochemical processes. Or, to put the point slightly differently, it remains to be shown why neurochemical (or quantum, for that matter) explanations don't threaten the causal efficacy of neuropsychological states if, as proponents of the exclusion argument claim, neuropsychological explanations threaten the causal efficacy of mental states. If each explanatory level undermines the levels above it, then the only real explanations are those of fundamental physics, and we're *all* out of a job.

It seems to me that there something particularly puzzling about *Wegner* suggesting that mental causation is merely virtual, for his own work on thought suppression and mental control highlights just how potent (and surprising) mental causation can be (see Wegner 1989). Indeed, Wegner's model of the conscious will is itself committed to the causal efficacy of mental content, for according to the model it is a match between the content of a person's intentions and their actions which generates the experience of doing.

¹³ See Heil and Mele (1995) and Walter and Heckmann (2003) for recent collections on the problem(s) of mental causation.

6. Conclusion

It is time to pull some of these threads together and draw some general conclusions. In order to know whether the experience of conscious will is an illusion we need an account of the content of the phenomenology of agency and an account of the structure of agency itself. We are still some way short of achieving either of these goals. This is perhaps most obvious with respect to the phenomenology of agency. As I suggested at the outset of this chapter, the content of the “experience of conscious will”—if we are to take this expression in its widest sense to refer to the phenomenology of first-person agency in general—is complex, vastly more complex than is usually assumed within the cognitive sciences. It may well turn out that some components of the phenomenology of agency are systematically misleading and other components are highly reliable. It may even turn out that some components—such as the experience of libertarian free will, if there is such a thing—are logically incapable of being satisfied. The recent literature on the phenomenology of agency has brought into focus some of the crucial issues that need to be addressed here, but it has also demonstrated just how much work remains to be done before we know, in the relevant sense, what it is like to be an agent.

We are also some way short of an adequate account of agency itself. Even so basic a distinction as the distinction between controlled and automatic actions is obscure and contested, with various theorists drawing the distinction in importantly different ways. And, of course, the role of the self and intentional states in the genesis of agency is even less clear. We will not know how much of the phenomenology of agency might be illusory until we know what role intentional states have in the explanation of human behavior. Despite these lacunae in our understanding it seems to me highly unlikely that the phenomenology of agency is systematically misleading. We experience ourselves as agents who do things for reasons, and there is little serious reason to suppose that we are mistaken on either count.

References

- Bayne, T.J. 2002. Review of D. M. Wegner, *The Illusion of Conscious Will*, *Journal of Consciousness Studies*, 9/7: 94-96.
- Bayne, T.J. & Levy, N. In Press. ‘The Feeling of Doing: Deconstructing the Phenomenology of Agency’, In N. Sebanz and W. Prinz (eds.), *Disorders of Volition*. Cambridge, MA: MIT Press.
- Bishop, J. 1983. Agent Causation. *Mind*, 92: 61-79.
- Block, N. 2003. Do Causal Powers Drain Away? *Philosophy and Phenomenological Research*, 67/1: 110-127.
- Felce, D. 1994. Facilitated communication: Results from a number of recently published evaluations. *Mental Handicap* 22: 122-26.
- Frith, C. 2002. Attention to action and awareness of other minds. *Consciousness and Cognition*, 11: 481-87.
- Ginet, C. 1990. *On Action*. Cambridge: Cambridge University Press.

- Haggard, P. in press. Conscious intention and the sense of agency. In N. Sebanz & Prinz (eds) *Disorders of Volition*. MIT Press.
- Haggard, P. & Johnson, H. 2003. Experiences of Voluntary Action. *Journal of Consciousness Studies*, 10, 9-10: 72-84.
- Haggard, P. & Libet, B. 2001. Conscious Intention and Brain Activity. *Journal of Consciousness Studies*, 8/11: 47-63.
- Haggard, P. Cartledge, P., Dafydd, M., Oakley, D.A. 2004. Anomalous control: when 'free-will' is not conscious, *Consciousness and Cognition*, 13: 646-54.
- Heil, J. & Mele, A. (eds) 1995. *Mental Causation*. Oxford: OUP.
- Hohwy, J. (in press) The experience of mental causation. *Behavior and Philosophy*.
- Holton, R. 2004. Review of "The Illusion of Conscious Will", *Mind*, 113/449:218-221.
- Horgan, T., Tienson, J., & Graham, G. 2003. The Phenomenology of First-Person Agency, In S. Walter and H-D Heckmann (eds) *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. Exeter, UK: Imprint Academic (pp. 323-40).
- Jacobson, J.W., Mulick, J.A., Schwartz, A.A. 1995. A history of facilitated communication: Science, pseudoscience, and antiscience. *American Psychologist* 50: 750-65.
- Jahanshahi, M. and Frith, C.D. 1998. Willed action and its impairments, *Cognitive Neuropsychology*, 15 (6/7/8): 483-533.
- Jeannerod, M. and Pacherie, E. 2004. Agency, Simulation and Self-identification, *Mind and Language*, 19/2: 113-146.
- Kim, J. 1998 *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.
- Metzinger, T. Conscious volition and mental representation: Towards a more fine-grained analysis. In N. Sebanz and W. Prinz (eds.), *Disorders of Volition*. Cambridge, MA: MIT Press.
- Nahmias, E. 2002. When consciousness matters: a critical review of Daniel Wegner's *The illusion of conscious will*, *Philosophical Psychology*, 15/4: 527-41.
- Nahmias, E., Morris, S. Nadelhoffer, T. & Turner, J. 2004. The Phenomenology of Free Will, *Journal of Consciousness Studies*, 11/7-8: 162-79.
- Nichols, S. 2004. The Folk Psychology of Free Will: Fits and Starts, *Mind and Language*. 19, 473-502.
- O'Connor, T. 1995. Agent Causation. *Agents, Causes, Events*. T. O'Connor (ed). New York: OUP: 173-200.
- Pacherie, E. 2000. The Content of Intentions, *Mind & Language*, 15/4: 400-432
- Peacocke, C. 2003. Action: Awareness, Ownership, and Knowledge. In J. Roessler & N. Eilan (eds) *Agency and Self-Awareness*. Oxford: OUP.
- Penfield, W. 1975. *The Mystery of Mind*. Princeton: Princeton University Press.

- Pockett, S. 2004. Does consciousness cause behaviour? *Journal of Consciousness Studies* 11, 23 - 40
- Preston, J., & Wegner, D.M. 2005. Ideal agency: On perceiving the self as an origin of action. In A. Tesser, J. Wood, & D. Stapel (Eds.) *On building, defending, and regulating the self* (pp. 103-125). Philadelphia, PA: Psychology Press.
- Wakefield, J. & Dreyfus, H. 1991. Intentionality and the Phenomenology of Action. In E. Lepore and R. van Gulick (eds) *John Searle and His Critics*. Oxford: Blackwell (259-70).
- Wegner, D.M. 1989. *White bears and other unwanted thoughts: Suppression, obsession, and the psychology of mental control*. New York: Viking/Penguin.
- Wegner, D.M. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
- Wegner, D.M. 2003a. The Mind's Self-Portrait. *Annals of the New York Academy of Sciences*, 1001: 1-14.
- Wegner, D.M. 2003b. The Mind's Best Trick: How We Experience Conscious Will. *Trends in Cognitive Science*, 7/2: 65-69.
- Wegner, D. M. 2005. Who is the controller of controlled processes? In R. Hassin, J.S. Uleman, & J.A. Bargh (Eds.) *The New Unconscious* (pp. 19-36). New York: Oxford University Press.
- Wegner, D.M., Fuller, V.A. and Sparrow, B. 2003. Clever hands: Uncontrolled intelligence in facilitated communication. *Journal of Personality and Social Psychology*, 85: 5-19.
- Wegner, D.M., Sparrow, B. & Winerman, L. 2004. Vicarious Agency: Experiencing Control Over the Movements of Others. *Journal of Personality and Social Psychology*, 86/6: 838-48.
- Wegner, D.M. and Wheatley, T. 1999. Apparent mental causation: Sources of the experience of will. *American Psychologist* 54: 480-491.
- Searle, J. 1983. *Intentionality*. Cambridge: Cambridge University Press.
- Searle, J. 2001. *Rationality in Action*. Cambridge, MA: MIT Press.
- Zhu, J. 2004. Is conscious will an illusion? *Disputatio*, 16: 59-70.