



Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Consciousness and Cognition 14 (2005) 771–785

Consciousness
and
Cognition

www.elsevier.com/locate/concog

Agency, authorship, and illusion

Eddy Nahmias

Georgia State University, Department of Philosophy, Atlanta, GA 30302–4089, USA

Received 30 April 2005

Available online 22 September 2005

Abstract

Daniel Wegner argues that conscious will is an illusion. I examine the adequacy of his theory of apparent mental causation and whether, if accurate, it suggests that our experience of agency and authorship should be considered illusory. I examine various interpretations of this claim and raise problems for each interpretation. I also distinguish between the experiences of agency and authorship.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Agency; Action; Free will; Phenomenology; Wegner

1. Introduction

Daniel Wegner suggests that our experience of consciously willing our actions is an illusion.¹ This claim has important implications for our sense of ourselves as agents and as authors of our actions. At a minimum, it suggests that things are not as they seem regarding that aspect of the world most important to us, ourselves. It might, for instance, imply that we are not free and morally responsible in the way we think we are. As Wegner puts it, “The fact is, it seems to each of us that we have conscious will. It seems we have selves. It seems we have minds. It seems

E-mail address: enahmias@gsu.edu.

¹ See Wegner (2005, 2004), Wegner, Sparrow, and Winerman (2004), Wegner and Wheatley (1999), and especially Wegner (2002).

we are agents. It seems we cause what we do. . . It is sobering and ultimately accurate to call all this an illusion” (2002, pp. 341–342).

But to determine if these claims are accurate requires, first, some analysis of what the experience of willing an action is, as well as its relation to the experiences of agency and authorship, and second, some analysis of what it means to say that an experience is illusory. Each of these tasks is monumental, so I will have to be somewhat sketchy. After examining Wegner’s ‘theory of apparent mental causation,’ I will argue that this theory gives us no reason to conclude that our relevant experiences of agency and authorship are illusory. More specifically, I will argue that: (1) having illusory experiences is not the same as having mistaken theories, (2) it would take a different sort of evidence than Wegner presents to show our experience of agency is illusory, and (3) the experience of *agency* or of willing an action is importantly different than the experience of *authorship* (or being the source of one’s actions), so that in whatever sense the former *may* be shown to be mistaken, the latter need not be. Though my discussion will focus on Wegner’s views, they may be applicable to other views that have a similar structure.²

2. The theory of apparent mental causation

Daniel Wegner’s experiments are ingenious and the unusual phenomena he describes in his work raise important questions about the psychology of agency, including the sorely neglected topic of the phenomenology of agency. In general, Wegner focuses on two types of situations that suggest there are notable exceptions to the ostensible rule that our voluntary actions are caused by our conscious intentions to perform them. First, there are situations in which people perform an action that looks voluntary but they do not experience themselves, or their own thoughts, to be the cause of the action. Examples include automatisms, alien hand syndrome, facilitated communication, and schizophrenia’s alien control. Conversely, there are situations in which people experience relevant thoughts about performing an action but the action is *not* in fact caused by those thoughts. Wegner has developed experiments that induce in his subjects an enhanced sense of agency and authorship for actions they do not in fact bring about (or even perform) simply because they are prompted to have a conscious thought that corresponds with an observed action and that occurs just prior to it.

For instance, in the ‘helping hands’ experiment, subjects look in a mirror at the movements of a confederate’s arms which are placed under the subjects’ arms to look like their own. When subjects hear a command for a certain type of arm movement (e.g., ‘make the OK sign’) that occurs just prior to their seeing the arm of the confederate making that movement, the subjects report an increased sense of controlling or willing the action compared to those cases when the command subjects hear does *not* match the movement they see in the mirror (or that comes well before the movement). That is, some sense of agency or authorship can be induced for actions the subjects clearly did not perform.³

² See, for instance, Bargh (2005) and Prinz (2003), who writes, “There appears to be no support for the folk psychology notion that the act follows the will, in the sense that physical action is caused by mental events that precede them” (p. 26).

³ See Wegner et al. (2004). See also the ‘I-Spy’ experiment in Wegner and Wheatley (1999).

Such experiments and phenomena led Wegner to the conclusion that our experience of mental causation derives from an unconscious inferential mechanism. Using Hume's account of our perception of causation in general, Wegner suggests that we experience a conscious thought (usually an intention) to be the cause of an action to the extent that the content of the thought is *consistent* with the action and immediately *precedes* the action, and we perceive no other likely cause for the action (*exclusivity*). This is the theory of 'apparent mental causation'—the idea that when our experiences satisfy these three conditions we unconsciously infer that our conscious thoughts caused our actions. As Wegner puts it, "the feeling that we consciously will our own actions is traceable to an inference we make from the match between our conscious thoughts and observed action" (Wegner et al., 2004, p. 839). Elsewhere, Wegner stresses that this inference issues in a "cognitive feeling" or "authorship emotion" (2004, p. 658) such that we feel like our conscious intentions cause our actions.

Below I will examine how one might interpret this theory to imply that our experiences are thereby *illusory*, which will require clarifying both what our experiences of agency are and what it means for an experience to be illusory. First, however, let us look more closely at the three conditions—consistency, priority, and exclusivity—Wegner associates with our experiences of agency.

It is unlikely that these three conditions will generally be *sufficient* for us to experience our normal sense of agency, at least for bodily actions. In addition, other factors, such as proprioceptive and kinesthetic feedback, normally play an important role in our experience of full-fledged agency.⁴ It is not just that our thoughts match our observed actions that produces an experience of agency, but we also have sensorimotor experiences of our body and its movements, and these experiences are part of a feedback loop between predicted and perceived movements which, if disrupted, diminishes one's experience of agency.

Sacks (1970) discusses the case of Christina, a woman who has lost her sense of proprioception. Eventually she learns to get around by coordinating her intentions with her visually observed movements—her intentions are then consistent with, temporally prior to, and perceived to be the exclusive causes of her actions. Yet, despite meeting these conditions of Wegner's model, she does *not* feel her normal sense of agency or self nor move as fluidly or efficiently. As she puts it, "It's like something has been scooped right out of me, right at the centre . . . see Chris, the first pithed human being [with] no sense of herself—disembodied Chris, the pithed girl!" (pp. 51–52). Christina's loss of the experience of her own body seems to disrupt her experience of herself as the source of her actions, even though she has the relevant experiences of priority, consistency, and exclusivity sufficient to *infer* that her intentions are the cause of her bodily movements.

It is likely that the missing proprioceptive feedback in Wegner's helping hands experiment also helps to explain why these subjects do *not* in fact report feeling that they controlled the relevant movements or acted intentionally in making the movements. Rather, they report a statistically significant higher degree of 'vicarious control' when the consistency and priority conditions are met, but the absolute value of the increase is only about 1 point on a 7 point scale, with the reports remaining on average below 3 out of 7. That is, subjects did report that they experienced a slightly enhanced sense of control when priority and consistency were present, but *not* that they experi-

⁴ Wegner discusses some of these other factors in Wegner et al. (2004, pp. 838–839). The case of *mental* actions, such as calculating, deciding, or planning do not involve proprioception, so Wegner's three conditions are more likely to be sufficient in these cases. I discuss mental actions more fully in Section 5.

enced themselves as controlling the action. Of course, in this experiment the exclusivity condition is not met. In the ‘I-Spy’ experiment, where the exclusivity condition was made ambiguous, satisfying the priority and consistency conditions also raised subjects’ judgments of ‘personal intention’ for causing an action, but only from about 45% to about 60% (see Wegner & Wheatley, 1999). That is, in these experiments Wegner is *not* in fact producing situations where subjects feel their actions are caused by their conscious intentions and demonstrating that this is not the case. Subjects do *not* report “having performed the movement intentionally” and it has not been shown that “the experience of will can be created by the manipulation of thought and action in accord with the principle of priority” (2002, p. 78).⁵

I also doubt Wegner’s three conditions are *necessary* for us to feel like we are the authors of our actions. We often experience ourselves as the authors of our well-rehearsed automatic behaviors, as in athletic or musical performances, even when we experience no immediately preceding conscious thoughts (e.g., intentions) to perform the specific behaviors involved. Rather, we seem to have a general intention or plan to play well, perhaps to carry out some array of actions, and then we let our bodies take over, consciously monitoring our movements but not consciously forming intentions to carry out specific subsequent actions (indeed, forming such conscious intentions tends to trip us up). Nonetheless, we experience ourselves as the source of the resulting actions. Perhaps in certain cases of well-rehearsed actions we may experience a *reduced* sense of agency; actions that require an exercise of ‘willpower’ or concentration may be accompanied by a greater sense of personal agency than skilled actions performed more effortlessly.⁶ Nonetheless, unless the priority condition is interpreted to include a much longer span of time than the 1–5 s Wegner suggests, it does not seem necessary for us to experience ourselves as the authors of our actions. I’ll return to this point in Section 5.⁷

Having raised these questions, I suggest that Wegner is nevertheless correct that the three conditions cited by his theory of apparent mental causation are surely significant *contributors* to our

⁵ It would be interesting to develop a ‘helping hands’ experiment that *did* create a sense in subjects of being the *exclusive* cause of the hands’ movements. Here’s a possibility (though it may not pass the Human Subjects’ Committee). First, numb subjects’ arms, but let them see that they can still move their arms without proprioceptive feedback. Then, surreptitiously place them in a set-up like the helping hands scenario. Then, have a well-disguised confederate’s arms carry out movements that match commands the subjects are asked to carry out (while the subjects’ numbed arms are restrained without their feeling it). If the set-up is convincing, these subjects would, I predict, report they had moved their own arms. However, without proprioceptive feedback, I suspect they, like Christina, would *not* report a full-fledged feeling of agency, and it may be that their reports are based less on a normal experience of agency than on an inference that they must have moved their arms based on the exclusivity principle: “Who else could have moved my arm?” Wegner suggests, rather, that the normal experience of agency is always based on such a causal inference that produces a cognitive emotion of conscious will.

⁶ This is one place where the phenomenology is complex and under-explored. Some cases of skilled behavior seem to be accompanied by a *heightened* sense of one’s self being engaged in and in control of the activity, while others seem to involve a diminished sense of agency (perhaps including ‘flow’ experiences).

⁷ Coordinated actions, as in ballroom dancing, may also raise questions about the relation between the experience of agency and the exclusivity condition. There are real-life examples of situations like the helping hands scenario, in which people have thoughts consistent with and just prior to observed actions caused by another agent. I have in mind synchronized swimmers or ballet dancers. It would be interesting to probe the phenomenology of such agents to see if satisfying the same two conditions as in the helping hands experiment produces any enhanced sense of control or agency for their partners’ actions.

normal experience of agency, and he demonstrates as much with numerous experiments that show that manipulating the presence of these conditions does influence subjects' reports about the *degree* to which they experience control or authorship for their actions. For now, let us assume that the theory of apparent mental causation is a roughly accurate account of the inference mechanisms that produce our sense of consciously willing our actions. The question then is why this would then lead to the further and more disconcerting conclusion that our experience of agency is *illusory*.

3. Illusory experiences

Wegner recognizes that his use of the word 'illusion' to describe our experiences of agency, of authorship—of having a self at all—is controversial and would provoke criticism (see his 2004, p. 682). A legitimate criticism is that he never clarifies exactly what he means by calling these experiences an illusion or what he means by 'the experience of conscious will.'⁸ For instance, he concludes his book by saying that it is an illusion "that we cause what we do" and "Our sense of being a conscious agent who does things comes at the cost of being technically wrong all the time" (2002, p. 342), but elsewhere he writes, "Questions of whether thought actually does cause action, for example, have been left in peace, and the role of consciousness in the causation of action has been ignored as well" (2005, p. 32). I will try to clarify the situation by explaining what I take it to mean for an experience to be illusory and then offering various ways of interpreting Wegner's claim that the experience of conscious will is illusory.⁹

An experience is illusory if the way it represents things is not the way things actually are. More precisely, a person's experience of X is illusory if the content of the experience includes that X has various features, but X does not in fact have those features. So, to show that an experience of X is illusory requires (1) showing that the content of the experience includes that X has certain features and (2) showing that X does *not* in fact have those features. It may also require (3) showing that even after the person comes to *believe* that X does not have the relevant features, she nevertheless continues to experience X as having those features (i.e., illusions are 'cognitively impenetrable').¹⁰ I will focus on the first two features for now.

⁸ Note that, assuming Wegner thinks that, by definition, our experiences are conscious, the phrase 'the experience of conscious will' is itself very confusing. Is it that we are conscious of being conscious of willing (presumably a rare occurrence) or is there just an extra 'conscious' used so that it should read 'the experience of willing [or intending] an action'? And, as I will discuss below, the experience of willing an action may be importantly different from the (presumably more general) experiences of agency or of authorship.

⁹ For further discussion of various interpretations of Wegner's theses, see Nahmias (2002), Bayne (Forthcoming), and commentaries in Wegner (2004).

¹⁰ Take, for example, our visual experience of Müller-Lyer lines: our experience includes the feature that the line with inward-pointing arrowheads is longer than the line with outward-pointing arrowheads, but the lines in fact do not have this feature, since they are equal in length. For most people the lines continue to appear unequal even after they come to believe they are equal in length. The most extreme cases of illusion involve experiences of something existing that does not actually exist (e.g., a hallucination). For agency and intention to be an illusion in this sense would require that agents and intentions do not exist at all. Wegner sometimes suggests this eliminativist position—e.g., "To be accurate, we must speak of apparent mental causation, or of virtual agency, rather than of intention or of a controller" (2005, p. 32).

So, what is the experience that Wegner is claiming is illusory—what is the ‘X’ and what features does it *seem* to have that it does not *actually* have? To answer this question, it is just as important to understand the relevant phenomenology as it is to understand the relevant psychological and neurobiological facts. And the relevant phenomenology has not been adequately explored, by Wegner or anyone else.¹¹ It likely involves various experiences, including experiences of *mental* actions such as deliberating and making decisions, as well as various experiences of *bodily* actions, including acting ‘thoughtfully’ (with one’s intention in mind), acting ‘thoughtlessly’ (without concurrent awareness of one’s intention), exerting oneself in action, feeling in control, and acting directly in response to ‘triggering’ stimuli (e.g., returning a shot in tennis).

Though all of these experiences likely involve the agent’s feeling that she is importantly involved in the action, it is far from obvious that they all involve the agent’s feeling that her *conscious* thoughts are causing her bodily movements. Wegner seems to have in mind those cases where an agent is consciously performing some action, and he takes the paradigmatic cases to involve ‘willpower’ and the “exercise of self-control” (2005, p. 30). So, the experiences in question involve an agent’s being aware of her intention to perform an action and being aware of performing that action. What then are the relevant *features* of these experiences that are supposed to be illusory?

Notice that the more substantial the purported experience (the ‘thicker’ the phenomenology), the more plausible it will turn out to be illusory, and conversely, the less substantial the experience (the ‘thinner’ the phenomenology), the less likely it will be illusory. More, and more detailed, features of experience mean there’s more to get wrong. Hence, it would advance Wegner’s conclusion if he could offer evidence that our experience of agency includes features that are unlikely (even impossible) to be true. Though, as I have suggested, he does not offer much evidence regarding the phenomenology, he may be read as suggesting our experience of agency includes a robust experience of a self, or homunculus, that is not mechanistic, suggestive of a Cartesian soul, and that directly causes our actions, perhaps even in an agent-causal way.¹² If our experience of agency is this robust, then Wegner is right that it is illusory. But I do not think that our experience includes, or commits us to believing in, anything as metaphysically sophisticated as a dualistic or agent-causal self.¹³ In any case, if *this* is the illusion Wegner has in mind, then: (a) he would not be offering any new revelations and (b) the sort of evidence he presents would not be particularly relevant.

A slightly less robust, though still Cartesian, conception of the experience of agency includes the feature of being ‘self-knowing’ or ‘self-luminous’ (2004, p. 682)—that is, in consciously performing an action, the agent is directly aware of the *cause* of the action, specifically, that the agent or his intentions cause the action. Then, the claim is that this experience is illusory because we do *not* have direct access to the causes of our actions and we can be mistaken about them. To show that our experience is illusory in this sense, it is not enough to show that we are sometimes, or even often, mistaken about why we do what we do—a fact well-confirmed by some of Wegner’s work, along with a long tradition of research in social psychology. It must also be established that we experience ourselves as being *infallible* about our own reasons for action, that the experience of

¹¹ For some recent discussions, see Bayne and Levy (Forthcoming), Horgan, Tienson, and Graham (2003), Nahmias, Morris, Nadelhoffer, and Turner (2004).

¹² See Bayne (Forthcoming).

¹³ See Nahmias et al. (2004) for discussion of the phenomenology of free will, including whether there is evidence that ordinary people have an experience of agent causation. See also Horgan et al. (2003).

action essentially includes awareness of the motivating causes of the action. I find it highly dubious that our experience of agency includes this experience of infallibility. Though we are likely overconfident about how much self-knowledge we have, we certainly seem to recognize what our friends, Freud, and a long literary tradition tell us—that sometimes we are wrong about why we do what we do—that sometimes our true intentions are not (easily) accessible to us. That we sometimes have to confabulate explanations for our actions (and sometimes know we have confabulated) offers some evidence that *during* action we do not have a ‘self-luminous’ experience of mental causation.

Perhaps the illusion, then, derives not from the purported experience of infallibility but from the purported experience of *direct access* to the causes of our actions. That is, when we *do* feel we know why we are acting, we are conscious of our relevant mental states as the cause of our actions, but these mental states are not the causes of our action. As above, we might question whether we do experience our mental states as causing our actions.¹⁴ I am inclined to say we do. But then the burden falls on Wegner to show that our mental states do not cause our actions. Sometimes Wegner suggests that our experience of agency is an illusion because we experience our conscious thoughts as the cause of our actions but we do not—perhaps cannot—experience the low-level mechanisms actually involved in producing both the thoughts and the actions. For instance, he writes, “The real causal sequence underlying human behavior involves a massively complicated set of mechanisms” (2002, p. 27), and “We should be surprised, after all, if cognitive creatures with our demonstrably fallible self-insight were capable of perceiving the deepest mechanisms of our own minds” (2003, p. 68).

But here it is more appropriate to describe our experiences as *incomplete* rather than illusory. We are quite lucky we do not experience most of what happens before and after we form our intentions, including all the neural work that goes on between the intention and the muscle movements, and also any unconscious inferential processes that inform our causal judgments. But this lack of direct introspection of what happens below the surface does not entail that our experiences are *illusory* any more than our lack of direct perception of the molecular activity that happens below the surface when one billiard ball strikes another entails that our experience of causation in that case is illusory. Rather, our experiences would be illusory in both cases only if there were no significant relationship (perhaps supervenience or identity) between the low-level processes and the interactions that occur at the level we *do* perceive. Unless, that is, we are willing to say that *all* of our experiences of causal interactions are illusory because we do not perceive the underlying causal mechanisms.¹⁵

Here is another way to put the point. Suppose the theory of apparent mental causation is accurate in that our experience of ourselves or our intentions as the cause of our actions is inferential. The fact that we experience such causation as *non-inferential* does not show the experience to be illusory unless the “conclusion” of the inference is systematically mistaken. So, it would have to be shown not merely that we do not know everything about the process by which we experience

¹⁴ See Horgan et al. (2003, p. 225).

¹⁵ It is somewhat ironic that Wegner cites Hume on causation since Hume *is* a skeptic about all causal relations (i.e., about our knowledge of any necessary relations between events), whereas Wegner is a skeptic about conscious will (mental causation) precisely because we do not experience the low-level mechanistic processes he dubs the ‘real causes of human action’ (2002, p. 97). A consistent Humean about causation should be no less skeptical about mental causation than any other type of causation, so long as there are the requisite observed regularities (see below).

our actions as caused by our conscious intentions but that, in general, the intentions (or other relevant mental states) that we are (sometimes) conscious of are *not* the causes of our actions. And, as I will discuss below in Section 4, this has not been established by Wegner nor could it be established by the sort of evidence he presents.

Perhaps, however, Wegner is better interpreted as suggesting not so much that our *experiences* of will and agency are mistaken, but that our *beliefs* about them are mistaken. That is, we may be radically *mistaken* about how the relevant causal processes work. Indeed, we may have *theories* about how causation, whether physical or psychological, works that are just dead wrong. For instance, we may have a mistaken theory about a force called ‘impetus’ that moves from one ball to the other when they strike. And it is quite possible that lots of people have false *theories* about what happens when they act—they may, for instance, have a dualist *theory* of mind and believe that their intentions are non-physical causes of their actions. But a false theory is not best described as an *illusion* in part because of the fact that false theories or beliefs, unlike illusory experiences, lose their hold on us when they are corrected (this was my third claim about illusions above).¹⁶ It should be no surprise that our folk psychological theories are not entirely correct, just as our folk physics and folk biology are not entirely correct. But these theories derive from more than our experiences, and our experiences are consistent with different (and more accurate) theories.

Sometimes Wegner seems to be suggesting this sort of approach to the psychological study of agency. That is, we should not take our experiences of agency or will to be any more indicative of the underlying nature and causal structure of ourselves, our minds, or our actions than we take our perceptual experiences to be indicative of the underlying nature and causal structure of the external world. For instance, Wegner writes, “The task of determining the causal relations between conscious representations and actions is a matter of inspection through scientific inquiry” because our experiences of such causal relations “can be misled by any number of circumstances that render invalid inferences” (2003, p. 68). This conclusion is an appropriate call for research. But who would disagree with it? Anyone studying the mind should be interested in the causal relations between consciously experienced mental states and bodily movements, and few cognitive scientists or philosophers would think that direct introspection is the best (or even primary) method for revealing those *causal* relations.¹⁷ Any physicalist about the mind will also seek to understand the neural processes that are the subvenient base of our conscious mental states and the causal relations between these processes and bodily movements.¹⁸

Indeed, any satisfying *physicalist* theory of mind will have to explain our conscious experiences, including for instance, why we experience our intentional actions at the level of purposeful mental states rather than mechanistic interactions. But if the idea is that the existence of such explanations shows our experiences to be illusory, it is too hasty, because our experiences alone

¹⁶ See Jack and Robbins commentary on Wegner (2004).

¹⁷ Even Cartesian dualists should want to study the causal interactions they think hold between (immaterial) conscious thoughts and bodily movements and understand when they are inaccurately perceived, though the difficulty of understanding what these mind–body causal interactions might be is a notorious problem for such dualists.

¹⁸ Again, if Wegner’s claim that conscious will is an illusion is just the claim that Cartesian interactionism is false, then it seems he presents the wrong kind of evidence. The right kind of evidence would be neurobiological evidence offering complete causal explanations of bodily movements by physical states (though this evidence would not be conclusive since overdetermination is a logical possibility).

do not commit us to the belief that such explanations do not exist. Similarly, our ordinary perceptual experiences are not illusory just because there are complex neurobiological explanations for those experiences. Our *experiences* of agency are best viewed as theory-neutral. They are consistent with a variety of theories about the metaphysics of mental causation, including some forms of physicalism. Though some people's experience of agency may incline them towards a dualistic theory of mind, a committed physicalist does not stop experiencing herself as consciously willing her actions.¹⁹

Instead, for our experiences of agency to be an *illusion*, they must involve a genuinely misleading experience of what actually happens when we act, a misleading experience that can be explained away but cannot be corrected in the way a theory can. For this to be the case, it would require not just that our experiences are incomplete or our theories about mental causation are mistaken, but that our experiences are systematically misinformative. What would it mean for our experiences of agency and authorship to be misleading in this way?

4. A one-way exit?

Wegner thinks that we experience our conscious intentions as the cause of our actions but they are *not* in fact the cause. Again, this does not follow from the fact that conscious intentions are not the *direct* cause of bodily movements without intervening mechanisms or the fact that conscious intentions may be physically instantiated. Rather, our experiences of conscious will would be systematically misleading in this sense only if there were some (set of) causes for our behaviors that systematically bypass conscious intentions entirely.

Assuming a physicalist theory, the only way I can make sense of this claim is that the neural processes involved in our forming conscious intentions are not causally connected to our action-production system—those neural processes involved in causing behavior. Wegner suggests this picture with his diagram of 'apparent mental causation' that has unconscious mental events producing both actions and thoughts about them, but the thoughts have no causal effects on the actions (there are no "actual causal path" arrows leading away from the "thought" box).²⁰ But Wegner's evidence for this view is not based on neurobiological discoveries that the two systems are in fact disconnected or that the system involving conscious intentions does not feed into the action-production system.²¹ Rather, he seems to base his claim primarily on the fact that sometimes we experience an intention without the appropriate action following, and sometimes we per-

¹⁹ Of course, it is an empirical question whether most people in fact have a dualistic theory of the metaphysics of mental causation. I suspect that these theories are influenced more by people's culture and religion than by their phenomenology of agency.

²⁰ See diagram in Wegner (2002, p. 68, or 2003, p. 66).

²¹ Wegner does discuss Benjamin Libet's experiments that purportedly show that actions are produced by brain activity that precedes conscious awareness of the intention to act, but there are important problems with interpreting Libet's work. For instance, Libet's data may only show that our *urge* to move precedes our awareness of the intention to move. Libet's data do *not* establish that the brain activity involved in our conscious intentions is not causally related to our actions. Wegner also references Penfield's work and theory of mind research to suggest that the intention-formation system may be a distinct 'module' from the action-production system (2002, chaps. 1–2), but even if they are distinct, that would not show that they are causally disconnected. See Nahmias (2002).

form what looks like voluntary movements without experiencing the appropriate intention. These exceptional cases are then taken to prove the rule that conscious intentions do not play any role in causing actions.

But of course the fact that events of type *X* sometimes occur without events of type *Y*, and vice versa, does not entail that when such events—in this case a conscious intention and the corresponding action—*do* occur together the *X*-type event is not the cause of *Y*-type event. All you need is the possibility of alternative effects of *X*-type events and of alternative causes of *Y*-type events for these exceptions to be possible and explicable.²² For example, our experience of seeing an orange may be caused by a hologram (and, of course, a real orange may not cause an experience of seeing an orange), but these exceptions do not entail that the presence of an orange is not usually the cause of our experiences of seeing an orange. Similarly, my experience of willing to move my arm may not cause my anesthetized arm to move, and my arm's moving as I dribble a basketball may not be caused by *immediately* prior conscious intentions to move them, but these exceptions do not entail that my conscious intentions are *never* the cause of my arm's moving.

I put this in terms of *types* because doing so allows us to see that there may be *token* instances of causal factors other than *X*-type ones that can bring about a token *Y*-type event. And vice versa. But again, these token instances may represent exceptions to the rule that *X*'s typically cause *Y*'s. For instance, in the extension of Wegner's helping hands experiment I describe in footnote⁵, we might induce in subjects a substantial sense of agency while we know that they are not causing any bodily movements at all. But in this case, as in *all* of the cases Wegner uses to show that we can experience some degree of agency without causing the relevant action, there is another causal source of the action as well as a causal explanation for the induced experience of agency. These exceptional cases certainly open up the *possibility* that our conscious intentions *never* play a role in action production, but they do not establish that possibility as actual.²³

Rather, the evidence required to establish this generalization would involve showing that the neural processes that cause voluntary behavior are not regularly influenced by the neural processes that are the subvenient base for (or are identical to) our consciously forming our intentions. The intention-system would be, as it were, on a one-way exit from the action-production system, as suggested by Wegner's diagram. While this sort of epiphenomenalism is possible, it is not suggested by the exceptional cases Wegner discusses—since, again, in these cases there is always some explanation for the disconnect between the experience of agency and the cause of action. On the contrary, numerous other experiments suggest that conscious intentions are causally implicated in

²² Alternatively, causal activity in one system *A* may normally be caused by causal activity in another system *B*, but sometimes similar activity in *A* might be caused by activity that bypasses *B*. This would not show that *B* activity is not the usual cause of *A* activity.

²³ Note also that the cases Wegner describes of an agent's actions that look voluntary but are not experienced as consciously intended (e.g., automatisms, alien hand) do not show that the agent has an *illusory* sense of agency, since the agent does *not* have a sense of agency in these cases. That is, these agents are not mistaken in believing that their conscious intention did not cause their action, because they did not *have* a conscious intention to perform the action. They are, of course, mistaken if they believe that they (i.e., their bodies and brains) are not causally responsible for their movements.

action.²⁴ Wegner’s work effectively explains *when* our experience of agency is an illusion but not that such experiences are *always* illusions.

One final reason to reject the view that our experience of conscious will is illusory is that these experiences are clearly functional. They allow us to act effectively in the world in the same sense that our visual experiences represent the world in a way that allows us to act effectively. Our visual experiences are, of course, *incomplete* in that they do not represent the micro-structure of the world but this tends to be a good thing as long as we are not doing physics (or perhaps metaphysics). Our visual experiences are also subject to illusions in that they sometimes misrepresent the *macro*-structure of the world. These illusions tend to occur, however, precisely because our visual systems evolved to represent the macro-world in a useful and accurate way. For instance, the environmental cues we use to perceive depth from a two-dimensional retinal image are the same cues that lead us to misperceive as unequal the equally long lines in the Müller–Lyer illusion and to misperceive a Necker cube as three dimensional.

Similarly, it should not be surprising that our experiences of willing our actions are *sometimes* subject to illusion in this sense. For instance, it is useful for us to interpret other agent’s intelligent behavior in terms of the agent’s beliefs and desires (this “theory of mind” system likely evolved in the context of our ancestors’ complex social environments). But this system may lead us to attribute mental states to complex systems (like computers or even the weather) that do not have such mental states. Such misattributions do not, however, entail that *no* agents actually have mental states that cause their actions. Wegner (2002, chaps. 6–7) discusses such misattributions (including anthropomorphism) and suggests that our attribution of conscious mental states as causes is always illusory but allows us to “construct a virtual agent in which we can reside” (p. 269). But just as our visual illusions occur against a background of functional perceptions, it may be that our mistaken attributions of mental causation do not represent a rule that no such mental causation exists but rather occur against the background of generally functional and accurate perceptions of our own and others’ actions being caused by the relevant mental states (some of which are conscious).²⁵

The claim that conscious will is an illusion has the ring of a skeptical argument. Such arguments have this basic form: we know our experiences are *sometimes* misleading, so we cannot know that they are not *always* misleading. Such arguments are notoriously challenging in the context of philosophical debates about epistemology.²⁶ But luckily, in the context of empirical psychology,

²⁴ Consider the interesting research by neurobiologists at Duke University led by Miguel Nicolelis (see, e.g., Carmena et al., 2003, and ‘Monkeys Consciously Control a Robot Arm Using Only Brain Signals’ at <http://news.mc.duke.edu/news/article.php?id=7100>). They record specific neural activity of monkeys moving a joystick so that a computer can read off such brain activity and then produce the corresponding movements before the monkeys do. The monkeys learn that they can simply form the intention to move and the computer will take over so that they do not have to carry out the movement of the joystick. Assuming the monkeys have conscious intentions (or that similar work could be done with humans), it appears that the neural processes involved in forming such intentions are causally implicated in subsequent actions.

²⁵ This is an appropriate place to refer the reader to Austin’s response (1977) to the so-called ‘argument from illusion,’ in which he points out that the existence of perceptual illusions that are virtually indistinguishable from veridical experiences (cases of which are much rarer than suggested by advocates of the argument) does *not* suggest that there are no important differences between them. There is usually a straightforward explanation for the unusual experience. Similarly, there are explanations (though perhaps less straightforward) for the unusual experiences Wegner creates in the lab or discusses.

²⁶ Challenging not because they force us to accept the conclusion but because they force us to analyze the concepts involved in the argument, such as “knowledge.” The arguments, of course, are more complex and complete than the one above.

we can look to empirical data to explore the likelihood of systematic illusion. In the case of perception, we can discover the mechanisms responsible for our misperceptions and explain why they depend on the mechanisms responsible for our normally perceiving the world accurately. In the case of intentional action, we might similarly discover the mechanisms responsible for the cases Wegner discusses and explain why they depend on the mechanisms responsible for our normally perceiving ourselves accurately. Of course, we may discover that we *are* systematically misperceiving the role of conscious intentions because they really are epiphenomenal—in the sense of being causally cut off from behavior production. But demonstrating this would require neurobiological evidence beyond what Wegner discusses. Discovering that the neural processes involved in the formation of conscious intentions do not causally influence the behavior-production system is possible but, I think, highly unlikely. Here's one reason why.

5. Agency vs. authorship

Suppose it did turn out that those conscious intentions that occur just prior to action are only effects, not causes, within the behavior-production system. Even this discovery would not show that conscious deliberation and intention formation are not causally implicated in our actions. This is because many of the intentions we form are not intentions to perform an action now but rather to perform various actions later. Here we must distinguish, first, *distal* intentions from *proximal* intentions (Mele, 1992) and, second, *conscious* intentions from *non-conscious* intentions. Distal intentions are intentions to perform an action at some future time, with varying degrees of specificity regarding the time to perform the action and how exactly to carry it out. And sometimes the intention is to perform an action in response to some predicted situation (or triggering stimulus). I may intend to go jogging sometime later today, or I may intend to offer a particular answer to the question I expect a student to ask in class. Distal intentions are often formed with conscious deliberation and we are usually conscious of them when we form them. However, we are often *not* conscious of these intentions just prior to acting on them. We are also not conscious of many of our *proximal* intentions, those that occur just prior to action. If it were shown that, in those cases when we *are* conscious of our proximal intentions, our being conscious of them is not causally relevant to the action, this would not in itself show that our consciousness of our distal intentions plays no causal role in our performing the actions later.

These distinctions allow us to recognize that even if it turns out that our experience of *agency* were shown to be illusory in the sense that our conscious experience of proximal intentions is causally irrelevant to our action, this would not thereby show that our experience of *authorship* is illusory. Here, it is important to be clear about the relevant phenomenology. It seems right to say that often the experience of *agency*—especially the cases of ‘willpower’ and self-control Wegner highlights—involves the feeling that one's conscious proximal intention to perform a specific movement is the cause of that movement. But the more general experience of being the *author* of our actions encompasses a much wider scope of intentions and actions. Reporting just my own phenomenology (I ask you to consult your own), I feel like I often carry out many complex actions without forming immediately prior conscious intentions to perform those acts. Rather, at some earlier time I form a general plan or a set of distal intentions to carry out an array of actions, and later (sometimes significantly later) I carry out those actions automatically without forming

conscious proximal intentions for each action.²⁷ I have in mind not just playing guitar or basketball ‘in the zone’ (unconsciously, as it were), but also the various actions following a decision to walk to the fridge to get a beer, to drive home, or even to deliver a well-planned lecture. In such cases, even if my conscious thoughts are not proximal causes of my actions, they seem to be crucial distal causes.²⁸

Wegner writes, “many of our most fluid, expert, and admirable acts are ones we do not experience consciously willing” and this leads to a “loss of the sense of authorship in skilled actions” which “do not feel willed as they unfold” (2005, p. 29). I do not think this is an accurate description of the relevant phenomenology. We do not experience our skilled actions as happening to us or as disconnected from our own intentions, plans, and goals, including some formed during conscious learning and conscious deliberation. Rather, we experience ourselves as the authors and sources of these actions, sometimes the more so precisely because of the earlier conscious effort we have put into them.

This explanation for our sense of authorship would offer an alternative explanation for certain cases of what Wegner calls “vicarious agency,” our experience of authorship for others’ actions. It is not that we feel “a twinge of authorship when our child wins an award . . . because our anticipatory thoughts of the glory made us vicarious agents in the action” (Wegner et al., 2004, p. 847). Rather, we may feel a twinge of authorship in such cases because we feel like we were important distal causes of our children’s success; we raised them to be so successful! Similarly, I can feel some sense of authorship for the eventual effects of some of my actions, even when my actions are not the proximal causes of those effects. Authorship, like responsibility, can be experienced to different degrees and can be shared among various agents.

Finally, Wegner has given us no reason to believe that our conscious thoughts are not causal factors in our *mental* actions, such as concentrating, deliberating, or making decisions. And it is these mental actions involved in deliberation and the formation of general plans of action that seem most significant to our sense of authorship and responsibility for our later actions, many of which may be carried out without any experience of forming a proximal intention to perform them. If you jump to the aid of a person in need or keep silent in response to someone’s degrading comment, you may feel responsible for your action or your omission even without forming conscious intentions to act (or not to act), because you may trace your response to earlier character-forming actions or to conscious considerations about what sort of person you want to be. We are authors of our actions in part because we are authors of our general plans and distal intentions.

It is *possible* that even these conscious deliberations leading to decisions and plans of action are just epiphenomenal effects of processes that could occur without input from the processes underlying conscious awareness, but I see no reason to think this is likely.²⁹ Rather, given the way the

²⁷ This is not to say that I do not have *intentions* to perform these actions since, as pointed out above, intentions can exist without our being conscious of them. It would be a mistake to say that we must be conscious of any intentions we have at the time we have them.

²⁸ Indeed, there is research on the ‘causal primacy effect’ to suggest that we generally attribute greater causality to earlier rather than later events in a causal chain, which would help to explain our experiences of authorship for actions caused in part by our prior intention formation.

²⁹ I am here ignoring the philosophical arguments that pose problems for the causal efficacy of conscious mental properties in general (e.g., Kim, 1998), since such possibilities neither support, nor are supported by, Wegner’s discussion. If the causal exclusion problem is the illusion Wegner is suggesting, then his evidence is not the sort needed to make his case. See Nahmias (2002).

brain is interconnected, it is unlikely that any significant set of neural processes—including those involved in conscious mental states—is causally cut off from those involved in producing behavior. Thus, even if more evidence turned up to show that we are not quite the *agents* we think we are—because our conscious mental states are not the immediate causes of our actions—we could still be the *authors* we think we are—our conscious mental states being significant distal causal contributors to at least many of our decisions and actions.

6. Conclusion: What am I afraid of?

In his recent reply to commentaries, Wegner (2004) scoffs at my whimsical use of Jerry Fodor's oft-cited remark that if our beliefs and desires do not cause our actions, then “practically everything I believe about anything is false and it's the end of the world” (1990, p. 196 quoted in Nahmias, 2002, p. 539). Wegner suggests that anyone who reacts with such “shrill invective” against his views must be motivated by emotion—“a motor, an emotional basis that drives the rhetoric” (2004, p. 680). This response is reminiscent of the defense a Freudian might use against his opponent: “Of course, you don't accept my theory. You're repressing your sexuality.”³⁰

Well, I am motivated (perhaps even by my emotions!) to try to understand and find fault with the claim that our sense of self, agency, and will are all illusory, because these are significant claims, ones that may have an influence beyond the ivory tower on people's conception of themselves and of ethical, legal, and political issues. (Of course, I may have other motivations that I do not even know about—my experience of agency is not infallible.) But let me be clear about what is *not* the “emotional basis” of my critique. I am not afraid of the fact that our conscious experiences fail to provide a direct view into the workings of our brain. Introspection thankfully does not allow us to experience the low-level neural activity involved in our coming to form our intentions and actions. Nor is introspection an infallible guide to why we do what we do or whether our conscious intentions cause our actions. Wegner is right to call for more psychological research into the phenomenology of agency and the causal mechanisms involved in action and the experience of agency. Nor am I afraid that our naïve theories of mental causation may be mistaken in significant ways. I am a physicalist. So if our folk psychology tends to be dualist, I think it *is* radically mistaken. But I do not think our experiences of agency would be illusory if dualism is false because I think our experiences are consistent with a physicalist ontology.

As a physicalist I am also not afraid that we are, in some sense, complicated mechanisms that can be studied scientifically, though I *am* motivated by the worry that it is very easy to misinterpret—and to *misrepresent*—this truth to suggest, for instance, that it shows that we are not really agents, that we do not really have minds, that we lack free will or moral responsibility, or that our sense of self is illusory. Future scientific study of the mind may show that some of these claims are accurate, but doing so will require that we analyze more carefully what our phenomenology of agency actually is, what it would mean for it to be illusory, and what the underlying neurobiology is. Indeed, I *am* afraid that such scientific study might someday show that our conscious deliberations, decisions, and intentions are *not* causally relevant to our actions (or even that they are

³⁰ Thanks to Neil Levy for this way of putting it.

much less relevant than we think—e.g., because they have only a retrospective role in helping us take responsibility for what our bodies do). But, luckily, Wegner has not shown this yet.³¹

References

- Austin, J. L. (1977). In G. J. Warnock (Ed.), *Sense and sensibilia*. New York: Oxford University Press.
- Bargh, J. (2005). Bypassing the will: Towards demystifying behavioral priming effects. In R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 37–58). New York: Oxford University Press.
- Bayne, T. (Forthcoming). Phenomenology and the feeling of doing: Wegner on the conscious will. In S. Pockett, W.P. Banks, & S. Gallagher (Eds.), *Does consciousness cause behavior? An investigation of the nature of volition*. Cambridge: MIT Press.
- Bayne, T., & Levy, N. (Forthcoming). The feeling of doing: Deconstructing the phenomenology of agency. In N. Sabanz & W. Prinz (Eds.), *Disorders of volition*. Cambridge: MIT Press.
- Carmenta, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., et al. (2003). Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biology*, *1*, 193–208.
- Horgan, T., Tienson, J., & Graham, G. (2003). The phenomenology of first-person agency. In S. Walter & H. D. Heckmann (Eds.), *Physicalism and mental causation: The metaphysics of mind and action*. Exeter: Imprint Academic.
- Kim, J. (1998). *Mind in a physical world*. Cambridge: MIT Press.
- Mele, A. (1992). *Springs of action: Understanding intentional behavior*. New York: Oxford University Press.
- Nahmias, E. (2002). When consciousness matters: A critical review of Daniel Wegner's. The illusion of Conscious will. *Philosophical Psychology*, *15*(4), 527–541.
- Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2004). The phenomenology of free will. *Journal of consciousness studies*, *11*(7–8), 162–179.
- Prinz, W. (2003). How do we know about our own actions? In S. Maasen, W. Prinz, & G. Roth (Eds.), *Voluntary action: Brains, minds, and sociality* (pp. 21–33). New York: Oxford University Press.
- Sacks, O. (1970/1987). *The man who mistook his wife for a hat*. New York: Harper & Row.
- 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. (2004). Précis of *The illusion of conscious will* [and Commentaries]. *Behavioral and Brain Sciences*, *27*(5), 634–692.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology*, *86*, 838–848.
- Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, *54*, 480–491.

³¹ Many thanks to the organizers of *The Brain and Its Self* conference at Washington University, St. Louis, April, 2004. I appreciate the invitation to participate and the useful discussion from the audience. I also appreciate the very helpful suggestions from two reviewers for this journal.