

When consciousness matters: a critical review of Daniel Wegner's *The illusion of conscious will*

EDDY NAHMIAS

ABSTRACT *In The illusion of conscious will, Daniel Wegner offers an exciting, informative, and potentially threatening treatise on the psychology of action. I offer several interpretations of the thesis that conscious will is an illusion. The one Wegner seems to suggest is “modular epiphenomenalism”: conscious experience of will is produced by a brain system distinct from the system that produces action; it interprets our behavior but does not, as it seems to us, cause it. I argue that the evidence Wegner presents to support this theory, though fascinating, is inconclusive and, in any case, he has not shown that conscious will does not play a crucial causal role in planning, forming intentions, etc. This theory’s potential blow to our self-conception turns out to be a glancing one.*

1. Introduction: a blow to our self-image

Freud claimed that his theory marked the third of the “major blows at the hands of science” to “the naïve self-love of men.” First, Copernicus moved us from the center of the universe, then Darwin knocked us from the pinnacle of creation, and finally, Freud showed us that “the ego is not even master of its own house, but must content itself with scanty information of what is going on unconsciously in its mind” (1989/1920, p. 353). But at least Freud offered *some* role for the conscious will, especially if, through psychoanalysis, we unveil the unconscious machinations that move us.

Daniel Wegner might be seen as a representative of a fourth scientific blow to our self-image, a post-Freudian knockout blow that says our conscious will plays no role in causing our actions but is, instead, an illusion. His *The illusion of conscious will* is an exciting, informative, entertaining, and potentially terrifying treatise on the psychology of action. In the end, Wegner seems uncertain about how threatening he means his view to be. He wants to expose as illusions what we take to be true: “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 we are agents. It seems we cause what we do” (p. 342). But he also wants to suggest that his explanation of how these illusions are

Eddy Nahmias, Department of Philosophy, Florida State University, Tallahassee, FL 32306–1500, USA, email: enahmias@fsu.edu

produced by the brain allows them an important place in our lives: “After all, the apparent mental causation approach does not throw out conscious will entirely; it instead explains how the experience comes to be ... Ultimately, our experience of conscious will may have more influence on our moral lives than does the truth of our behavior causation” (pp. 334, 341).

I don’t think he can have it both ways. If we accepted his theory as true, it would threaten our conception of ourselves as free and responsible (as Wegner recognizes in various places—e.g. in his contrast between the positive illusion of control and a “resignation to determinism”; pp. 332–333). In any case, I will argue that his thesis is overstated. Wegner has gathered a tremendous amount of interesting empirical work and presented it in an engaging manner. Despite his arsenal of empirical findings, however, he has not shown that our conscious will is an illusion—at least not in the strong sense that says our conscious experience of willing our actions plays no important causal role in how we act. Weaker interpretations of Wegner’s thesis, more accurate in light of the research he presents, suggest that this blow to our self-image is a glancing one.

2. Interpretations of *The illusion of conscious will*

Wegner summarizes his main thesis in several different ways, and it can be difficult to discern exactly what he is trying to say and, hence, how threatening his thesis is to our sense of self and free will. I will describe three ways one might interpret the claim that conscious will is an illusion. Wegner hints at the first two but argues most clearly for a third interpretation. I will argue that the evidence he marshals in support of this interpretation is unconvincing.

Wegner writes, “The experience of consciously willing an action is not a direct indication that the conscious thought has caused the action. Conscious will, viewed in this way, may be an extraordinary illusion indeed” (p. 2). What might this mean?

2.1. *Dualism is false?*

One way to interpret the claim that conscious will is an illusion is that the prevalent commonsense view of the mind—what Ryle called the “Official Doctrine”—is wrong: there is not a mental force that causes bodily behaviors, no ghostly soul that moves the machine. So, in the quotation above, the mistake would involve our Cartesian belief that “conscious thought” certainly doesn’t *seem* like a physical entity but it certainly *does* seem like it moves us to act. The illusion would be analogous to our old belief that the sun revolves around the earth because it certainly *seems* like we are not moving and the sun is.

It would be odd if dualism is the illusion Wegner is describing, since this claim would be uncontroversial to many of his readers, certainly to most psychologists and philosophers. Furthermore, though neuroscientific evidence continues to whittle down any role for a non-physical soul, this is not the sort of evidence Wegner discusses. Yet sometimes it seems he is trying to make a case against dualism when he emphasizes that the way mental causation *appears* to us is like Cartesian interac-

tionism—“We intended to do it, so we did it” (p. 27)—but the way actions are actually caused involves mechanisms about which we are unaware. Wegner writes, “people have at hand two radically different systems of explanation, one for minds and one for everything else” (p. 21). We apply the mentalistic explanations, with their reference to beliefs, desires, purposes, and intentions, to our own and others’ behaviors: “we find it enormously seductive to think of ourselves as having minds” (p. 26). But, Wegner suggests, such explanations are mistaken in the sense that “the real causal sequence underlying human behavior involves a massively complicated set of mechanisms ... [and] the mind can’t ever know itself well enough to be able to say what the causes of its actions are” (pp. 27–28).

Again, this claim is uncontroversial to anyone who thinks we are not conscious of the neural mechanisms in our brains. And the claim that we are not conscious of the causes of our behavior becomes tautological when he asserts, “The real causes of human action are unconscious” (p. 97). What *is* controversial is how to cash out these claims. If Wegner is not just denying the untenable dualism suggested by our ordinary language of mental causation, perhaps he is suggesting that conscious processes simply cannot cause physical behavior.

2.2. *Epiphenomenalism is true?*

Perhaps the way to interpret the illusion, then, is that all conscious mental states simply do not have any causal influence on what we do; they are epiphenomenal. The role of consciousness is illusory, then, because despite how it seems, the existence of conscious will has no effects on how we behave. It is a side effect, like Huxley’s example of the steam whistle that is caused by the train’s engine but does not move the train (cited on p. 317). Wegner suggests that the conscious will is caused by the “engine” of the brain (the “empirical will”) but does not itself move us to act: “Whatever empirical will there is rumbling along in the engine room—an actual relation between thought and action—might in fact be inscrutable to the driver of the machine (the mind)” (p. 28). Perhaps conscious will could even be stripped off of the mind’s activity without this affecting what we do. This strong view is related to the issues philosophers worry about when they discuss the problem of zombies, absent qualia, and of course, mental causation. One worry is that, if the cause of behavior can be sufficiently explained by low-level (“unconscious”) mechanisms, what role is left for conscious processes?

In fact, I don’t think Wegner means to raise this metaphysical worry, despite his suggestions that mechanistic explanations are real and mentalistic explanations are illusory. Though his bibliography includes an impressive number of references to works of philosophy, he does not really engage with the philosophical questions about mental causation. Instead, he denies that conscious will is entirely inefficacious: “the experience of conscious will that is created in this way [by unconscious mechanisms] need not be a mere epiphenomenon. Rather than a ghost in the machine, the experience of will is a feeling that helps us to appreciate and remember our authorship of the things our minds and bodies do” (p. ix; see Chapter 9). Nonetheless, he clearly has in mind some sort of epiphenomenalism: “Why, if

this experience of will is not the cause of action, would we even go to the trouble of having it? What good is an epiphenomenon?" (p. 318). So what does he have in mind?

2.3. *Modular epiphenomenalism is true*

The third way to interpret the claim that conscious will is an illusion puts an empirical face on the metaphysical claim of epiphenomenalism. It says that the conscious will is explicable in terms of brain processes but that these brain processes are not "in the loop" when it comes to producing actions. It is *not* that conscious mental states in general are epiphenomenal but that specifically those thoughts and intentions we experience just before actions as the cause of those actions do not in fact cause our actions. Furthermore, the conscious experience of will is not entirely impotent—it "helps us appreciate and remember what we are doing" (p. 325), and it is "the person's guide to his or her own moral responsibility for action" (p. 341). (In offering these roles for conscious will, Wegner thinks he can dampen the blow of his thesis and even provide an explanation of free will—but see discussion below.)

Wegner thus suggests what I will call "modular epiphenomenalism." He describes conscious will as a module or system in the brain that functions to interpret behavior in terms of mental states such as beliefs, desires, intentions, and plans. But this module, responsible for producing the experience of will, is separate from the modules or systems that actually cause behavior, what he calls the "empirical will." Most of the evidence discussed in the book is used by Wegner to demonstrate that the experience of will and the empirical will "come apart often enough to make one wonder whether they may be produced by separate systems in the mind" (p. 11). The "interpretive module" is epiphenomenal, then, in the sense that its activity is independent from and occurs too late to affect the brain systems that cause behavior. Its function is to observe and theorize about behavior, not produce it.

So, modular epiphenomenalism is an empirical claim about the timing of and connections between events in the brain: sluggish conscious processes are too slow to occur before behavior is generated, and they are on a one-way exit from the behavior-production system: "The brain, in turn, shows evidence that the motor structures underlying action are distinct from the structures that allow the experience of will. The experience of will may be manufactured by the interconnected operation of multiple brain systems, and these do not seem to be the same as the systems that yield action" (p. 49) [1]. The brain systems involved in producing conscious intentions are modules (or, when diagrammed, boxes) which receive input about what behaviors are being produced but which themselves produce no output relevant to the behavior they represent. Wegner's central diagram captures this idea.

Wegner cites three sources of evidence to support the anatomical and temporal claims of this modular theory. First, he discusses psychologists, such as Alan Leslie and Simon Baron-Cohen, who argue that "theory of mind" (our ability to understand mental states and their relations to behavior) may be a specialized module (p. 24). This view of theory of mind is controversial, but even if true, it would not show that theory of mind always involves retrospective interpretation, never prospec-

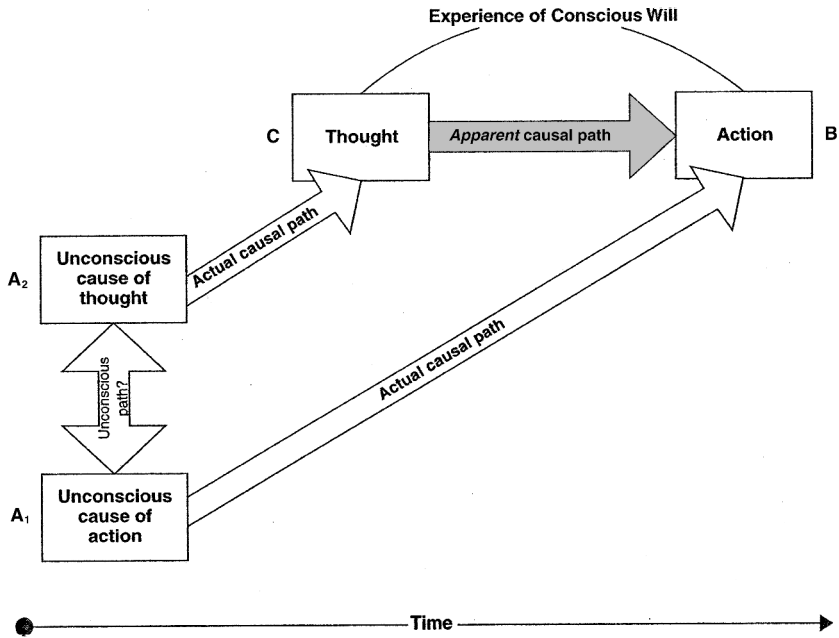


FIG. 1: The experience of conscious will arises when a person infers an apparent causal path from thought to action. The actual causal paths are not present in the person's consciousness. The thought is caused by unconscious mental events, and the action is caused by unconscious mental events, and these unconscious mental events may also be linked to each other directly or through yet other mental or brain processes. The will is experienced as a result of what is apparent, though, not what is real. *Source:* Wegner (2002, p. 68).

tive influence. Many actions involving, for instance, social coordination, deception, and self-control require theory of mind (as suggested by the difficulty they pose for autistics, young children, and animals which supposedly lack theory of mind). There is no reason to think this module (if it is a module) represents “the endpoint of a very elaborate inference system underlying apparent mental causation” (p. 325) rather than an important component in planning and executing some types of action.

Wegner's second source of evidence for modular epiphenomenalism is drawn from Benjamin Libet's experiments on the role of conscious will in voluntary behavior. These well-known experiments suggest that unconscious neural activity (the readiness potential) precedes subjects' conscious awareness of their desire to, say, move a finger. Wegner uses these results to suggest that the experience of conscious will might not even be a link in the “causal chain leading to action”; instead, “it might just be a loose end—one of those things, like the action, that is caused by prior brain and mental events” (p. 55). The image, which recurs throughout the book, is that unconscious mental activity *A* is a common cause for both behavior *B* and the experience of willing the action *C*: *A* is the *real* cause of *B*, and *C* is a separate effect that comes too late to influence *B*. (Wegner also notes that there may be two separate unconscious causes *A*₁ for *B* and *A*₂ for *C* and that *A*₁ and *A*₂ may or may not be connected—see Figure 1.)

But Libet himself suggests that our awareness of volition occurs in time to veto the action—so conscious will is not causally cut off from action production. More important considerations involve common critiques of Libet’s work. First of all, Libet’s discussion neglects the role of earlier conscious intentions—for instance, the subjects’ accepting the instructions and deciding to spontaneously lift a finger later. Libet’s results, unless generalized as Wegner proposes, do not challenge the causal role of the “action plan” (or pre-intention) formed when subjects agree to do what is asked of them. Second, there is the difficulty of picking out the phenomenology and timing of “willing,” and of trying to time it precisely. Try spontaneously lifting your finger several times (and pay attention to the experience). When I ask people to do this, they report quite different feelings: some feel themselves initiate the movement, some feel the urge come upon them, and some feel the desire to move only *after* they move. The language Libet uses in various places (“urge,” “desire,” “decision,” “intention”) is as diverse as the phenomenology. In any case, as commentators have often observed (see Libet, 1985), anyone who is not a dualist should not be surprised that brain activity precedes conscious awareness. The question is whether, as Wegner suggests, *only* unconscious processes cause action and all brain processes “underlying” conscious will (to use Wegner’s vague language of the metaphysical relation) are epiphenomenal. Libet’s results do not show this.

Finally, Wegner argues that the modular view is supported by experiments using direct electrical stimulation to different parts of the cortex. Some stimulations produce behaviors experienced as unwilled—“I didn’t do that. You did” (p. 45)—while others produce behavior that is accompanied by an experience of will (so subjects then invent a reason why they performed the stimulated behavior). Wegner concludes: “the brain structure that provides the experience of will is *separate* from the brain source of action” (p. 47). This interpretation neglects the possibility that different stimulations may simply hit the complicated “causal chain” in various places—after all, you can hit my knee with a mallet and I’ll move it without experiencing the movement as willed. In any case, it should not be surprising that the brain’s systems for perception of action and production of action are not identical. We need to perceive actions without producing them when, for instance, we observe others acting, and we need to produce actions without perceiving what we are doing when we act very quickly (e.g., hitting a fastball). But to say the systems are not identical is not to say they are unconnected [2].

3. Exceptions as rules?

Wegner seems to realize the tenuous nature of the support these research programs provide for his conclusions. Thus, he often uses tentative language to describe his conclusions: “This, in turn, suggests the interesting *possibility* that conscious will is an add-on, an experience that has its own origins and consequences. The experience of will *may not be* very firmly connected to the processes that produce action, in that whatever creates the experience of will *may* function in a way that is only loosely coupled with the mechanisms that yield action” (p. 47, my italics). There is nothing wrong with scientists being cautious in their interpretations of empirical work—it

should be commended. But Wegner's use of such language can itself be interpreted in two ways—as an indication of his caution or as an indication that the work he cites can only support the claim that *sometimes* the experience of will can occur without the relevant action and vice versa.

In fact, most of the experiments and experiences Wegner discusses throughout the rest of the book are best seen as supporting this interpretation—that there are various exceptions to the rule that our conscious experiences of our actions correspond with those actions. But the fact that there are these exceptions does not show that, in the *normal* cases of correspondence, conscious will is causally irrelevant. Just because consciousness of willing an action *can* be separated from the action does not mean that when they are not separated, consciousness is not causally relevant (just because *A can* occur without *B* does not mean that when *A does* occur with *B*, *A* is not a cause of *B*) [3].

Wegner asserts that the exceptions may in fact be the rule, but he offers little support for that claim beyond the hypothesis of modular epiphenomenalism itself [4]. For instance, after discussing automatisms—behaviors an agent performs without experiencing conscious will, such as moving a Ouija board or automatic writing—Wegner concludes, “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: automatism is the rule, and the illusion of conscious will is the exception” (p. 143). His idea is that automatisms are produced by “the same kinds of processes that create intentional action” but the interpretive process, an add-on for actions experienced as intentional, is bypassed or obstructed in some way so it does not produce the illusion of conscious control (p. 130). But another interpretation is that automatisms are *not* produced by the same kinds of processes that create intentional action precisely because the causal role of conscious intention has been bypassed. It seems that neuroscientific data would be most useful to test between these hypotheses, but Wegner offers no such evidence (beyond that discussed above, which does not directly apply to automatisms).

Wegner argues in Chapter 3 that the interpretive process that creates the experience of conscious will works according to the theory of “apparent mental causation.” We experience our conscious thoughts and intentions as causing our actions when they fit three criteria which, he claims, drawing on Hume, “people use in the perception of causality more generally” (p. 64): (1) *priority*—the conscious thoughts and intentions (hereafter, just “thoughts”) must precede the behavior they represent, usually by 1–5 seconds; (2) *consistency*—the thoughts must be consistent with the actions (this will often be apparent because they are semantically related); (3) *exclusivity*—we believe our thoughts cause our actions to the extent that we do not perceive other factors that might be likely causes of the actions [5].

Most of Wegner's examples focus on cases in which one of these three principles breaks down so that we do not experience ourselves as causing behavior that we are in fact causing (illusion of non-control) and cases in which, because the three principles are met, we experience ourselves as causing behavior that we are *not*

in fact causing (illusion of control). For instance, in the “I Spy” study, by creating the experience of consistency, exclusivity, and especially priority, Wegner could induce subjects to think they controlled a computer mouse when they did not (pp. 74–78). Conversely, people feel out of control when they perceive some alternative cause of their behavior (thus losing the experience of exclusivity), such as internal impulses or external agents (as in hypnosis or schizophrenics’ hearing voices). Wegner cites the classic automatism (Ouija boards, automatic writing, etc.) as prime examples of voluntary actions—though the use of “voluntary” seems questionable—which are not experienced as consciously willed, usually because the principle of consistency breaks down. We do not understand how our specific movements could be caused by our thoughts since we do not consciously intend those movements. Sometimes our actions are even produced by trying to resist them, as in ironic processes in which trying *not* to think about or do something can increase the likelihood of thinking about or doing it (Wegner is known for his work on ironic processes, summarized in *White bears and other unwanted thoughts*, 1989).

Other examples Wegner discusses in depth (in Chapter 4) include dissociative experiences, ideomotor action, dowsing, and the chevreul pendulum, each of which involves agents’ producing movements without feeling like they are. In Chapter 8 he analyzes hypnosis in a similar way: the empirical will succumbs to the suggestions of the hypnotist while the conscious will, “associated specifically with an executive or controlling module of mind,” is bypassed (p. 301). The hypnotized agent does not feel out of control but he does not feel he is consciously controlling his actions either. In Chapters 6–7 Wegner discusses the converse cases in which our interpretive module goes awry and attributes conscious intentions and control where there is none. This so-called “action projection” includes the case of Clever Hans (and other examples of anthropomorphizing animals and objects), the tragic case of facilitated communication [6], and joint ventures in which people may project their intentions onto those of the group. It also includes the creation of “virtual agents,” such as ghosts, spirits, gods, and alternate personalities, by projecting purposeful intentions onto patterns of events or behavior. Wegner ends the chapter by postulating that these exceptional cases may represent the rule for how individuals create a sense of self: “We achieve the fact of having a perspective and being a conscious agent by appreciating the general idea of agents overall and then by constructing a virtual agent in which we can reside” (p. 269).

I should point out that Wegner describes all of these cases with lucid prose and charming wit. Most of the examples are inherently interesting since they are, well, weird, but Wegner also slips in lots of jokes and puns to keep the reader entertained (the bad puns are better than no puns at all). I certainly recommend the book as an informative and fascinating journey through some of psychology’s puzzles—and as I explain in my conclusion, I recommend it as a type of book that needs to be reckoned with. But returning to his central thesis, I think Wegner has misinterpreted, or at least exaggerated, the ramifications of these psychological puzzles.

4. When consciousness matters

When all is said and done, Wegner has offered no evidence or arguments against this proposal: certain brain processes have the property of being consciously represented to the agent as the mental states we describe as beliefs, desires, intentions, and actions (for instance, my brain is currently going through processes which I experience as something like “I think this proposal makes sense,” “Type out the words ‘this proposal makes sense’,” and so on). How it is that these brain processes have these experiential properties is currently a mystery (that is the hard problem of consciousness). But if these processes did not have their representational properties, then they would not have the causal powers they have (for instance, they would not allow me to represent various ways of carrying out my intentions and hence would not allow me as much flexibility to inhibit or adjust my actions) [7]. Hence, my conscious experiences have important causal influences on my actions.

Granted, I do not represent the brain processes themselves and so, in some sense, my experience of conscious will is illusory (perhaps in the same sense that my experience of tables as solid is illusory). But my experience is not illusory in the sense of being epiphenomenal or even inessential for most of my actions. And even though the conscious experiences may involve brain systems distinct from the systems that produce some behavior, consciousness is not a separate module, a box which receives input but produces no output relevant to behavior. Furthermore, there is no reason to suggest that these brain processes have their causal powers only in virtue of their mechanistic, neural properties rather than their psychological/functional/representational properties—no more than there is reason to suggest that the neural properties “lose out” their causal powers to the quantum-level properties they are identical to (or supervene on), or that genes do not have causal properties because only DNA molecules do [8].

Finally, *sometimes* I behave in ways that look (but don’t feel) similar to purposeful actions but without engaging the brain processes that have these representational properties (e.g. hypnosis and automatism). And sometimes my behaviors, or those I observe in other agents, are so similar to behaviors that are usually accompanied by conscious intentions that I (unconsciously) attribute conscious control to them (e.g. anthropomorphism). That purposeful behaviors and conscious awareness of their causes can come apart should not be too surprising, but without the “rule” that they usually do *not* come apart, the exceptions would not be intriguing. Indeed, I think Wegner’s intriguing book has in fact offered some evidence that *supports* the proposal I have just described.

But even if Wegner’s modular epiphenomenalism is more accurate than my proposal, he does not give conscious experience its proper due. Perhaps to make his thesis more exciting by making it more threatening, he seems to push it towards the more expansive epiphenomenalism that allows no important role for consciousness will in producing actions. He allows the experience of will a retrospective role—it connects us to our past actions—but he allows it no prospective role. This is a mistake.

There are several ways that conscious representations of future actions seem

essential to the way we perform them. One involves inhibition. For many types of action, in order to inhibit doing them, it is crucial to be able to represent them. Consider examples of intentional deception that require inhibiting an action (say, an eye gaze towards a hidden object) to avoid providing information to a target (theory of mind may also be necessary for such deception). Libet's interpretation of his experiments suggests conscious awareness of our volitions allows us to inhibit them (what he calls "veto power"). Wegner himself writes, "Conscious processes are more flexible and strategic" than unconscious ones (p. 57). This is in part because they allow us to represent and compare various actions and chose which of them satisfies certain goals. And ironically, the ironic processes Wegner studies suggest that conscious volition *is* causally relevant to action since, in these instances, trying to achieve a certain goal (e.g., to inhibit thoughts about white bears) often *causes* the goal to be frustrated. These exceptions, however, should not be confused with the rule that conscious intentions to achieve a goal generally facilitate actions that achieve the goal.

Indeed, the most obvious way that conscious will plays a causal role in behavior involves developing general goals or plans for behavior—that is, the activity involved in deliberating, deciding, forming prior intentions, and carrying out extended actions (see, for instance, Bratman, 1987; Searle, 2002). It is strange that Wegner discusses some of these points when he overviews the philosophy of action early in the book (pp. 18–20), only to ignore them when he advances his main thesis. His modular epiphenomenalism seems meant to apply to the conscious intentions we experience as occurring only just before we act, and most of his examples do not address planning, deliberation, or extended actions. But I think these are activities in which we feel our conscious thoughts and intentions as causally influencing our actions even if they occur much earlier than the actions (and even if we do not then experience an immediately prior conscious thought).

For instance, when we decide to drive home, to propose marriage, to reconcile with a friend, to study, or to water the plants, we form general "action plans" without filling in most of the detailed behavior that will constitute the actions. Indeed, a little phenomenological investigation seems to suggest that we do not consciously will most of our specific movements, except in cases where the actions are being learned, are difficult, or require inhibition, etc. The "quick actions," such as typing, that Wegner discusses (pp. 56–59) to demonstrate that consciousness is too slow to occur before such actions seem instead to bolster the idea that conscious will simply does its job earlier—for instance, when we decide what to type or that we will try to hit the fastball or, in the Libet experiment, that we will move our finger at some unspecified time. Wegner instead concludes that "consciousness pops in and out of the picture and doesn't really seem to do anything" (p. 59).

Furthermore, what we feel most responsible for are the plans and policies we make for our own lives, the decisions we make after conscious deliberation about what sort of motivations we want to move us, what types of actions we want to perform, what sort of person we want to be. Conversely, we can also feel responsible for actions we perform without conscious will. Consider an athlete who is playing "unconsciously"—she may be hyper-conscious of what is going on around her but

she is not consciously willing her specific actions. Nonetheless, she feels responsible because she has trained precisely so that she can act without consciously willing her actions [9]. Musicians and actors aim for the same goal. Indeed, we all aim to be able, in many types of situations, to act without the feeling of will, but in accord with the plans we have consciously formed. And we feel free and responsible for the actions that accord with these plans, regardless of whether we consciously will each of them.

Wegner ends his book with his own solution to the question of free will. Basically, he suggests that by explaining how the brain gives rise to the experience (the illusion) of conscious will, we can explain why we feel in control of some of our actions and feel morally responsible for them. This, of course, is no solution at all, since no one wants the experience of conscious will, self-control, and personal responsibility to be illusory, to be mere feelings that do not correspond to reality. What we want is that our conscious selves have a real influence on what we do, on what we become. We want (prospectively) to act on our reasons, not first to act and then (retrospectively) to rationalize why we acted. We want to know what we're doing and (at least sometimes) why we are doing it. I think we can have this sort of free will, and it is most evident in our conscious, sometimes self-reflective, deliberations and plans (see, for instance, Bratman, 1984; Nahmias, 2001; Searle, 2002).

There is a reason we don't think the correlation between a lifetime without stepping on a sidewalk crack and a mother never breaking her back is a real causal correlation. Even if we are too superstitious to test the counterfactual, we cannot imagine how the two events are physically connected in the right way. But the correlation between conscious intentions, plans, and thoughts and the actions they represent are not so physically isolated. Even if, as Wegner suggests, they are produced by distinct systems in the brain—a claim I have disputed—nothing in the brain is completely isolated. It would be a miracle if the system that deliberates, decides, and forms general action plans did not causally influence the system that produces bodily movements. Or perhaps it would just be epiphenomenalism—the strong kind that says no conscious process does anything. But since Wegner does not discuss what his modular epiphenomenalist thesis has to say about these types of conscious processes, he leaves the reader wondering how threatening the illusion of conscious will is meant to be [10].

5. Conclusion: why philosophers should respond to Wegner's work (and work like it)

Wegner claims, "Philosophers have given us plenty of 'isms' to use in describing the positions that can be taken on this question [of free will and determinism], meanwhile not really answering it in a satisfying way" (p. iv); they have "generated a literature that is immense, rich, and shocking in its inconclusiveness" (p. 26). We philosophers would respond that we have not solved this problem because it is not easy! In fact, some claim our limited human brains will never be able to understand how our brains are conscious or how it is that we could have free will (e.g., Noam

Chomsky's and Colin McGinn's mysterianism and Peter van Inwagen's recent views on the mystery of metaphysical freedom—positions I do *not* agree with).

But there is a danger in the way philosophers have dealt with the problem of free will. We have generated volumes in the deadlocked debates between compatibilists and incompatibilists, without appreciating that these two camps generally agree about *almost* everything necessary for free will—everything but whether indeterminism of some sort is essential. We generally agree, for instance, that free will requires that our conscious deliberations make a difference in what we do, that we have the ability to act according to our reflectively considered desires and reasons, and that we have conscious control over some of our actions as we perform them. So, when someone—like Wegner, who aptly recognizes the essential connection between the problems of consciousness and free will—suggests that we don't even have these abilities, we should collectively take notice.

We should take notice because other people will. For most philosophers the problem of free will may be about whether it is compatible with the metaphysical thesis of determinism. But the rest of the world does not feel as threatened by metaphysical vagaries (which may be unverifiable). For them, the problem of free will, when it goes beyond the sociopolitical problems of freedom, is usually about whether free will is compatible with the way science tells us we are (and I hasten to add that, despite confusion about the issue, scientific explanation is not equivalent to deterministic explanation). What makes people worry about free will is a type of reductionist view disseminated through popular science books, like Robert Wright's *The moral animal* and Matt Ridley's *genome*, and through ubiquitous media reports about genes for personality traits, brain areas that control deliberation and emotion, and evolutionary stories about our sexual, addictive, and aggressive instincts [11]. These views represent the scientific revolution that novelist Thomas Wolfe summed up for thousands of Duke graduates at the 2002 Commencement:

Man's view of himself is changing very rapidly. Within the next twenty years I think you'll feel its effects. After Darwinism, after Freudianism, there's now a tremendous influence of neuroscience ... If I may reduce with terrible reductiveness the bottom line of neuroscience, it is that let's not kid ourselves, we're all concatenations of molecules containing DNA, hard wired into a chemical analogue computer known as the human brain, which as software has a certain genetic code. And your idea that you have a soul or even a self, much less free will, is just an illusion.

The idea that the self and free will are illusions may never be able to force its way into our collective consciousness (though this will likely depend on what that idea is supposed to mean). But Wolfe's public address is buoyed by books like Wegner's, despite Wegner's concluding attempts to put a happy face on his thesis. And such views matter. The degree to which people feel they can control their actions and be responsible for them varies across cultures, eras, and individuals. Information from the sciences (and philosophy) can influence how these feelings play out, as demonstrated by debates over legal responsibility—from the insanity defense to the Twinkie defense—and political philosophies—from conservatives' view that individ-

uals are generally responsible for their position in life to liberals' view that society bears much responsibility for individuals' adversities.

In the face of these important issues, philosophers should not sequester their debates about free will to issues like metaphysical determinism, which may be relevant but whose relevance is often far from clear. We should engage with books like Wegner's because it is part of our job to examine the philosophical and ethical implications scientists draw from their experimental results and to make sure they are not inferring more than they should—indeed, a book like Wegner's helps remind a philosopher how important our distinctions and debates (e.g., about the mind–body problem) can be. At the same time, we should also learn about the empirical facts such books present so that our own theories are not out of touch with the way things are.

I have suggested that Wegner's *The illusion of conscious will* is successful in presenting a host of empirical facts that inform us about the way we think and act (though mostly in marginal situations). We should pay attention to these facts. But the book is not successful in presenting a decisive challenge to the folk intuition at the heart of philosophical conceptions of free will, that our conscious experiences of our deliberations, planning, intentions, and actions often play an essential role in what we do. Were it otherwise, then this fourth scientific revolution would truly be a knockout blow to what Freud called our “naïve self-love”—or, as Jerry Fodor put it, in the context of a closely related debate: “If it isn't literally true that my wanting is causally responsible for my reaching ... and my believing is causally responsible for my saying ... then practically everything I believe about anything is false and it's the end of the world” (1990, p. 196). We philosophers should keep our guards up against any blow that would be the end of the world [12].

Notes

- [1] Notice in this quotation (and others cited below) a phenomenon common throughout Wegner's book (and many scientific books that discuss the mind–brain relationship): the words used to describe the relationship—“underlying,” “allow,” “manufactured by,” “yield”—are vague and sometimes contradictory. Questions about the metaphysical relationship between mental and physical properties (e.g. supervenience, identity, causation, correspondence, etc.) are avoided.
- [2] Interesting research on “mirror neurons” has found neurons in primates' premotor cortex that activate both when the subject moves in particular ways and when the subject perceives similar movements in others, suggesting that at least in this brain area, representation of action production and action production are not cut off from each other, perhaps facilitating imitation (see Gallese *et al.*, 2002).
- [3] This sort of fallacious reasoning sometimes appears in discussions of phenomena like blindsight when people argue that, because (some) visual stimuli *can* be discriminated without conscious awareness, consciousness is not in principle necessary to visual perception. Or more comprehensively: because it seems possible to perform complex behaviors like ours without consciousness (e.g. robots), *our* conscious processes are not causally necessary for our behavior.
- [4] However, another source of evidence Wegner takes to support the rule that “consciousness doesn't know how conscious mental processes work” (p. 67) comes from social psychology research that suggests we cannot correctly introspect on the causes of our mental states or behaviors—e.g. generally we misattribute causation to character traits rather than situational factors. In Nahmias (forthcoming), I argue that the implications of such research have been

- exaggerated, often because, like Wegner, the theorists interpret abnormal cases as representing the norm.
- [5] Wegner's use of Hume is suspect (see pp. 13–14, 64, 66); he suggests that Hume showed us that our perception of the constant conjunction between our conscious intentions and our actions is not evidence of a real causal relation, but Wegner does not apply Hume's skepticism to the low-level mechanistic processes dubbed the "real causes of human action" (p. 97). Also, a consistent use of Hume would allow that, even if conscious will and action are not *constantly* conjoined, certain *instances* are; the exceptions might, as I suggest, be *exceptions*. Note also that Humean accounts of causation are controversial.
- [6] Facilitated communication involves facilitators holding their hands over those of a disabled person to help him or her type messages on a keyboard. Despite their belief that they are not active in composing the messages (their lack of conscious will), it has been shown that they are responsible for the content of the messages.
- [7] I present this counterfactual in terms of physical (or natural) necessity: I make no claim about consciousness being metaphysically (much less, logically) necessary for certain behaviors. For the purposes of understanding Wegner's empirical claims, this type of necessity is the relevant one.
- [8] This analogy between different examples of reductionism is, of course, too simple, but the point is that Wegner's continual claim that only unconscious mechanisms are real causes requires an argument he does not provide.
- [9] Olympic soccer star Tiffany Milbrett explained her game-tying goal during stoppage time of the 2000 gold-medal match in this way: "I had no thoughts. My body just did what it did." Also, see the Norman–Shallice (1986) model of action which posits two complementary processes in controlling and guiding action, one that deals with novel or difficult actions that require conscious attention and one that deals with simple, well-learned, and habitual skills that do not require attention. The latter involve "motor schemata" which can carry out quite complex movements, though consciousness of the behavior still allows the agent to become aware of any deviations and correct for them (as when a musician hits a wrong note).
- [10] The closest Wegner comes to discussing the role of consciousness in deliberation and planning is when he mentions controlled mental processes and actions (p. 97), where he says, "The ability to know what one will do, and particularly to communicate this to others verbally, would seem to be an important human asset," but he also claims that "even if [these actions] occur along with an experience of control or conscious will, this experience is not a direct indication of their real causal influence." He refers to a forthcoming article where he will spell out this argument, "Who is the controller of controlled processes?" My guess is he will end up sinking into the stronger form of epiphenomenalism.
- [11] See Chapter 17, "Blaming the victim," of *The moral animal* (Pantheon, 1994), and Chapter 22, "Free will," of *genome* (Fourth Estate, 2000). Earlier books that sparked these flames include Richard Dawkins' *The selfish gene* and E.O. Wilson's *On human nature*. Needless to say, each of these books has been more widely read and discussed than any recent philosophical work on free will.
- [12] Thank you to Al Mele, Brandon Towl, and Cheryl Kopec Nahmias for helpful comments. Funding support from an FSU FYAP grant.

References

- BRATMAN, M.E. (1987). *Intentions, plans, and practical reason*. Cambridge, MA: Harvard University Press.
- FODOR, J. (1990). *A theory of content and other essays*. Cambridge, MA: MIT Press.
- FREUD, S. (1989/1920). *Introductory lectures on psycho-analysis*. New York: Norton.
- GALLESE, V., PEIRFRANSCESCO, F., KOHLER, E. & FOGASSI, L. (2002). The eyes, the hand, and the mind: behavioral and neurophysiological aspects of social cognition. In M. BEKOFF, C. ALLEN & G. BURGHARDT (Eds) *The cognitive animal*. Cambridge, MA: MIT Press.

- LIBET, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8, 529–566.
- NAHMIA, E. (2001). Free will and the knowledge condition. Unpublished dissertation.
- NAHMIA, E. (forthcoming). Free will and the threat of social psychology.
- NORMAN, D.A. & SHALLICE, T. (1986). Attention to action: willed and automatic control of behavior. In R.J. DAVIDSON, G.E. SCHWARTZ & D. SHAPIRO (Eds) *Consciousness and self-regulation: advances in research and theory*, 4, 1–18.
- SEARLE, J. (2002). *Rationality in action*. Cambridge, MA: MIT Press.
- WEGNER, D.M. (1989). *White bears and other unwanted thoughts*. New York: Viking Penguin.
- WEGNER, D.M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.