HOW TO ACCEPT WEGNER’S ILLUSION OF CONSCIOUS WILL AND STILL DEFEND MORAL RESPONSIBILITY

Richard Double
Edinboro University of Pennsylvania

ABSTRACT: In The Illusion of Conscious Will, Daniel Wegner (2002) argues that our commonsense belief that our conscious choices cause our voluntary actions is mistaken. Wegner cites experimental results that suggest that brain processes initiate our actions before we become consciously aware of our choices, showing that we are systematically wrong in thinking that we consciously cause our actions. Wegner’s view leads him to conclude, among other things, that moral responsibility does not exist. In this article I propose some ways that traditional philosophical defenders of moral responsibility, both compatibilists and libertarians, might accept Wegner’s empirical premise regarding the will but amend their theories so that they may reject his conclusion regarding moral responsibility.

Key words: conscious will, moral responsibility, readiness potential, action, compatibilism, incompatibilism, libertarianism, epiphenomenalism

Introduction

In this article I sketch a strategy by which philosophical critics might accept Daniel Wegner’s (2002) empirical thesis that he calls the “illusion of conscious will” (ICW) while rejecting the denial of moral responsibility Wegner draws from it. In the first two sections of this article I summarize Wegner’s account of human action and his conclusion regarding moral responsibility. In the rest of the article I offer three moves familiar to philosophers to show how we can give up our natural commitment to the efficacy of our choosing our actions and still defend moral responsibility.1 I see the three replies I shall offer as logically distinct, but as working best when taken together. My maneuver has this qualification: If our actions are as drastically disconnected from our consciousness as Wegner argues

AUTHOR’S NOTE: I am grateful to Edmund Abegg, Michael Long, Saul Smilansky, Terry L. Smith, Bruce Waller, Daniel Wegner, two anonymous referees of this journal, and especially the editor for helpful commentary on this article. Please address all correspondence to Dr. Richard Double, Philosophy Department, Edinboro University of Pennsylvania, Edinboro, PA 16444. Email: rdouble@edinboro.edu

1 For a broad range of responses to ICW from social scientists, neuroscientists, and others see Libet, et. al. (Eds.), (1999), which is a reprint of a 1999 special issue on the volitional brain in Journal of Consciousness Studies, 6, 8-9. For a sustained argument that Wegner does not establish ICW see Nahmias (2002).
that they may be, no modest adjustment of traditional philosophical themes such as those I suggest can save moral responsibility.

Wegner’s Empirical Premise About Consciousness of the Will

Wegner (2002) offers a temporal account also presented by Benjamin Libet (1999, 2002). Unconscious neural states begin our volitional actions (actions we perform on request) by going into characteristic readiness potentials (RPs) approximately 800 milliseconds before we perform those actions. A readiness potential may be understood as an initial excitation of that part of our brain that typically produces the action in question. Our conscious awareness of our choices then typically occurs 500 milliseconds after the start of the readiness potential. Thus, our brains initiate our actions before we are conscious of choosing them:

The conscious willing of finger movement occurred at a significant interval after the onset of RP. . . .These findings suggest that the brain starts doing something first. . . .then the person becomes conscious of wanting to do the action. This would be where the conscious will kicks in. . . .in the sense that the person first becomes conscious of trying to act. (p. 53)

What all this means is that consciousness is kind of a slug. . . .(C)onsciousness seems to follow these actions rather than lead them. (p. 58)

I believe that this argument for ICW is compelling if and only if Wegner adds three premises. First, that for each distinct type of action (e.g., moving my hand to the right) there is a distinctive readiness potential, for if one type of readiness potential could initiate both my moving my hand to the left and to the right, then an objector could assign conscious will the role of dictating which of the possible directions one moves one’s hand. Second, that conscious will lacks the efficacy to prevent the action initiated by the readiness potential. Libet (1999, 2002) differs from Wegner by imputing to conscious will a veto power to stop readiness potential-initiated actions from coming to fruition. Third, that this temporal phenomenon holds for all actions (i.e., those requiring deliberation and planning, not just simple actions).

Although ICW strikes me as decisive given these three premises—which I am prepared to grant—Wegner devotes much of his book to exotic and not-so-exotic examples that show how often we are wrong when we take ourselves to be aware of the causes of our behavior. Exotic cases include alien hand syndrome, “actions” performed with phantom limbs, dowsing, Ouija boards, spirit possession, table turning, hypnotism, automatisms (e.g., automatic writing, speaking in tongues, ideomotor action), dissociative identity disorder, and experiments in split-brain patients. Wegner also cites many experiments from cognitive psychology to show that we are typically wrong about the causes of our behavior. Examples are action projection (e.g., Clever Hans), cognitive dissonance reduction, priming effects, motivated and unmotivated cognitive errors, and self-fulfilling prophecies.
The two themes (the temporal argument and the many examples) mutually support ICW. Interestingly, Wegner seems to see ICW both in a radical way and in an extremely radical way. Less radically, Wegner might view ICW as refuting our everyday belief that we consciously cause our actions. Although ICW, taken less radically, is quite noteworthy, I shall argue that it leaves enough of our commonsense view of persons intact to permit the sort of defense of moral responsibility I offer in this article.

More radically, Wegner sometimes seems to view ICW as a datum to be explained by an even more daring hypothesis. Wegner suggests that conscious will is merely a fictional property our brains ascribe to fictional (“virtual”) agents, perhaps similar to the way that patients suffering from dissociative identity disorder construct their multiple personalities (Ch. 7). At his most radical, Wegner sees an analogy between persons being directed while in a hypnotic trance and the real condition of human mentality (Ch. 9). The hypnotist is the cause of the subject’s actions, as the unconscious neural states cause the actions of the unhypnotized human; the actions of both are unaffected by the conscious will of the subject. The difference is that when we are being led while in a hypnotic trance we realize that our conscious will is ineffective; the unhypnotized person also has an ineffective conscious will but does not realize it (i.e., he suffers the illusion of conscious control). So, humans who are led by hypnotists are instructive examples of conscious automata, which we all are.

Wegner’s Conclusion About Moral Responsibility

Wegner’s refutation of moral responsibility, accordingly, can take two forms. The least dramatic is based solely on ICW. Wegner holds that conventional wisdom regarding moral responsibility requires that our conscious will cause our actions:

A person is morally responsible only for actions that are consciously willed . . . If conscious will is illusory, how can we continue to hold people responsible for what they do. . .? How can we blame people for despicable acts, if they didn’t will them? (p. 334)

Libet agrees, “We do not hold people responsible for actions performed unconsciously, without the possibility of conscious control” (2002, p. 58). The more dramatic argument is based not only on ICW but also on the hypothesis that it is “ultimately accurate to call illusions” the belief that we have selves and minds, that we are agents, and that we cause what we do (pp. 341-342). If there are no agents, it seems that there is no one to be morally responsible. In Ch. 9 Wegner says that the illusion of conscious will is helpful, especially in creating a sense of ownership of the actions we perform, thereby augmenting the helpful illusion of moral responsibility. Although I believe there are good philosophical reasons for denying the existence of moral responsibility (Double, 1991, 1996), I think that ICW, in its less radical form, does not show this.
Philosophical Defenders of Moral Responsibility: Compatibilists and Libertarians

Most philosophers care about the free will problem because of their interest in moral responsibility in its nonbehavioral sense. On this traditional view, persons who are morally responsible for their actions are appropriately subject to negative sanctions, not on the forward-looking ground that their behavior can be modified to produce socially accepted behavior, but because they deserve those sanctions. The concept I recommend is: “deserving positive and negative reactive attitudes (gratitude, resentment), praise and blame, retribution and moral reward for one’s behavior.” Some philosophers, such as Daniel Dennett, and some social scientists minimize or leave out this just-deserts aspect of moral responsibility, focusing on the social role that holding persons responsible plays. They, thus, increase the likelihood that moral responsibility will seem compatible with determinism. In doing so, however, they ignore the role that holding ourselves morally responsible plays in the special ways we see ourselves (regarding guilt, blame, and moral censure). They also ignore the historical origins of the worry over moral responsibility, e.g., the question of the merit of theologically-inspired attempts to, in Leibniz’s terms, “vindicate God’s justice” in the punishment of sinners. It is, of course, one’s right to care about whatever one wants, but adopting a sense of “moral responsibility” that leaves out just-deserts means that someone else will have to address that issue.

At the cost of some oversimplification, there are two main schools of philosophical defenders of moral responsibility in the just-deserts sense described above. Compatibilists believe that persons can be morally responsible for determined actions, provided that they want to perform those actions (David Hume), they reflectively endorse their desires to perform those actions (Harry Frankfurt), or they are sensitive to good reasons for and against those actions (John Martin Fischer). Compatibilists believe that instances of determinism that facilitate subjects’ meeting these conditions contribute to moral responsibility, whereas instances of determinism that thwart these conditions reduce moral responsibility.

Libertarians, following Kant, view compatibilism as a “wretched subterfuge.” Libertarians are incompatibilists, who believe that no one can be morally responsible for a determined action, whether it meets the compatibilists’ criteria or not. Libertarians also hold that some actions are not determined, and that we can be morally responsible for those actions provided we meet certain intelligibility constraints. In sum, to sustain a just-deserts variety of moral responsibility, compatibilists believe that determined subjects must choose from reasons, whereas libertarians demand both that we choose from reasons and make undetermined choices. The challenge for compatibilists is to give a persuasive account of how we can be morally responsible for actions that are determined (hence, inevitable). Libertarians must show how we can be morally responsible for actions that are not determined (hence, arguably not under our control).

Compatibilists and libertarians, although disagreeing over whether responsible choices would be determined, assume what ICW challenges, namely
that our conscious will is causal. I shall argue that compatibilists and libertarians can surrender this assumption to accommodate Wegner’s empirical hypothesis while maintaining the central tenets of their theories.

**Reply 1: Conscious Reasons *qua* Distal Causes (or Influences)**

Reply 1 sets the stage for more nuanced varieties of compatibilism and libertarianism. I include the term “influences” so that libertarians as well as compatibilists may use this reply. Although libertarians believe that responsible actions cannot be causally determined, many are happy to say that our beliefs and desires “probabilistically cause,” “bring about,” or “influence” our actions.

To begin, notice that the first sentence does not entail the second:

(1) Due to the ICW, in principle our conscious willing of our actions cannot cause them.

(2) Our conscious will is completely wrong about what the causes of our actions are.

The gap between (1) and (2) provides room for a response from defenders of moral responsibility.

According to the temporal argument, conscious will cannot cause the action we think it does because conscious will occurs after the action is initiated—but if we accept ICW and agree that our conscious awareness does not reveal the cause of action A, one still wishes to ask *what went into the causes of A?* Wegner removed conscious willing from the role of the proximate cause of our actions, but he still owes us a substitute explanation of how actions come to be.

Wegner endorses cognitivist explanations of actions throughout his book. Such explanations cite folk psychological states such as beliefs, desires, plans, and goals that compatibilists use in their accounts of free will and moral responsibility. Wegner, thus, leaves room for slow, plodding conscious thoughts *indirectly* to cause or influence our actions. For example, the halting, laborious, self-conscious learning of a foreign language or golf swing can lead to automatic, unconscious mastery. We admire a batter who, through long training to recognize the spin of thrown baseballs, has acquired the reflex to swing at a low fastball in the strike zone but to avoid the slider that will land in the dirt. By analogy, citing one’s earlier conscious rehearsing of reasons for actions allows for a *distal cause* compatibilist view of moral responsibility (or a *distal influence* libertarian view), even if our *present* conscious awareness does not affect our actions proximately.

It is possible that the brain’s unconscious beliefs, desires, and values—psychological aspects of the engine that drives behavior—are *themselves* caused by previous psychological states that we have consciously examined in the past.\(^2\)

---

\(^2\) I assume throughout that it is an open question whether *mental states and events* such as conscious and unconscious beliefs and desires are physical (as materialists maintain) or
Nothing in ICW precludes such distal ancestors from being the ultimate sources of our present actions, thereby making us responsible for those actions.

Consider an example from Dennett (1984). Dennett is convinced that, if asked, he would refuse to murder for money, and hopes that this value is chronic rather than transitory. Although Dennett’s present conscious awareness of his principle not to murder for money, according to ICW, could not be the proximate cause of his declining to murder now, his value’s (conscious) ancestors may be the distal causes or influences that ultimately produce his refusal to murder now. So, Dennett’s earlier conscious thinking about his principle may be the remote cause of his present action—even supposing that he would be wrong to credit his action to his present conscious choice. Compatibilists (or libertarians) would have to admit that our phenomenological impression is wrong, but they would emphasize that our folk psychological states motivate our responsible behavior through its scientifically discovered origins, our readiness potentials.

In the process of revealing to us our ownership of our actions, which Wegner endorses (p. 317), conscious will may reveal to us the chronic dispositional states that caused our present actions. By doing so, conscious will may serve as feedback to modify future actions. Such information may prompt us to ask, “Was I right to choose as I did?” and “Do I wish to be the kind of person who chooses that way?” Once we raise these questions, our lumbering conscious deliberations need not be a façade of confabulations (although many will be). They might result in the resolve to perform different actions the next time, a resolve that affects future actions.

Now, this opening response may seem only to push back Wegner’s challenge to moral responsibility one step. Even if our distal conscious reflections cause (or influence) our unconscious will as I suggest, who is to say that those distal conscious reflections themselves were not the products of still earlier exercises of unconscious will? If they were, citing them will not disrupt Wegner’s picture; it will simply add another layer to it. The distal conscious reflections themselves would be as far removed from our rational control as our simple actions according to ICW, in which case Wegner’s argument against moral responsibility remains unscathed.

This sensible objection is similar to one often made against Frankfurt-style compatibilism. If determinism is true, then all our actions are caused, including our evaluations of our actions, our evaluations of our evaluations, and so on. By definition, if determinism is true then there is no escape from it. The case is likewise if Wegner’s ICW holds universally.

Although this point is correct (citing possible distal conscious causation by itself is not a full vindication of moral responsibility), it would be an error to dismiss the importance of distal conscious causation. Emphasizing the possibility that distal conscious states cause or influence choices whose readiness potentials begin before we are aware of them opens the door to defending moral nonphysical (as dualists and idealists maintain). The alternative to leaving the question open is to beg the question against one side or the other.
responsibility despite ICW. Libertarians and compatibilists need to add more specifics from their theories to strengthen the response begun in this section.

Reply 2: Teleological Intelligibility

In the mid-twentieth century, philosophers of history debated whether historians should model their explanations after those of the physical sciences, i.e., deducing specific events from premises citing general (“covering”) laws and the state of the world prior to those events (“initial conditions”). This was known as “deductive–nomological explanation,” championed by logical positivists such as Carl Hempel. Their opponents held that historical explanation should not follow the deductive model, but rather should aim to make historical events understandable by citing reasons. This narrational sort of explanation was known as “verstehen explanation,” after the German word for “understanding.” It found its most prominent proponent in William Drey. In the free will discussion, verstehen explanations have become known as “teleological intelligibility”—explanations that work in virtue of citing the goals of agents rather than deterministic laws.

I have argued elsewhere (Double, 1996) that libertarians are teleological intelligibility theorists in the following sense: as indeterminists who believe morally responsible behavior cannot be explained deterministically, libertarians believe that if one can tell a plausible narrative about libertarian agents that cites reasons they actually hold for opting for each of two actions, then we should regard them as choosing responsibly either way. Verstehen explanations, in terms of intelligible goals, show that undetermined actions may be enough under agents’ rational control, despite indeterminism, for agents to be morally responsible for them.

I believe that teleological intelligibility accounts are attractive to many libertarians because a good story can convince us that both outcomes accord with a chooser’s total set of reasons. To use Robert Kane’s example of the business woman rushing to an important meeting who witnesses a mugging (1996, pp. 126-127), we can empathize with her reasons that incline her to stop to help the victim (if she does), as well as her reasons that incline her to rush to her meeting (if she does that). Therefore, according to Kane, the woman is morally responsible for either action she performs indeterministically because whatever action occurred was her action, done for her reasons. For the teleological intelligibility theorist, a plausible narrative trumps the familiar objection that undetermined choices are necessarily out of our control. Most importantly, if we add to the libertarians’ teleological intelligibility doctrine the hypothesis that our conscious reflections are distal influences of our choices (Reply 1) we get a modified libertarian picture of moral responsibility that is not clearly weaker than the standard libertarian view. Distal influences make undetermined libertarian choices intelligible; thus, persons are morally responsible for the actions they produce.

Allowing myself some terminological license, compatibilism can be viewed as a type of teleological intelligibility theory by emphasizing how a good narrative
can overcome our fear of determinism. Imagine what teleological intelligibility compatibilism looks like if we try to make moral responsibility compatible with ICW. Before, when compatibilists had to put a happy gloss on determinism, they could do so by appealing to the efficacy of their reflective, conscious choices. In this article, however, I envision them as accepting ICW, so they have to adjust their strategy.

Admittedly, compatibilists look askance at teleological intelligibility explanations when used by libertarians to argue that persons can be morally responsible for two-way, undetermined choices. Such explanations in the hands of libertarians reject the need to explain why the choice that was made was chosen for the reasons it was chosen, rather than the other choice for the other reasons (i.e., they fail to provide contrastive explanations). So, to a compatibilist, libertarian teleological intelligibility narratives are inherently unexplanatory.

Things are different, however, when a determinist adopts teleological intelligibility. A determinist who believes in moral responsibility (i.e., a compatibilist) could find teleological intelligibility explanations sufficient for showing that we are responsible for one-way, unconsciously determined actions. If we accept ICW and admit that conscious choices may be distal causes of our present actions, we see that teleological intelligibility explanations, ceteris paribus, more plausibly support compatibilism in a deterministic world than they do libertarianism in an indeterministic world. On both Wegner’s view and the libertarians’ we lack antecedent (proximate) conscious control over our free actions. Compatibilists can say that so long as determined actions are intelligible in light of the agent’s total set of reasons, the agent is morally responsible for the action, irrespective of the lack of proximate conscious control. Compatibilists such as Guy Claxton (1999), who argued that both compatibilists and libertarians exaggerate the importance of the “supervisory I,” might be sympathetic to the teleological intelligibility response to ICW. Teleological intelligibility is natural for compatibilists who accept ICW.

We might embellish the compatibilists’ use of teleological intelligibility by emphasizing the idea of choosing in character. The great historical figures held less rarified views of free will than contemporary compatibilists. Thomas Hobbes (1588-1679) argues:

[H]e is free to do a thing, that may do it if he have the will to do it, and may forbear if he have the will to forbear. And yet if there be a necessity that he shall have the will to do it, the action is necessarily to follow; and if there be a necessity that he shall have the will to forbear, the forbearing also will be necessary. (1841, p. 42)

John Locke (1632-1704) expresses compatibilism this way:

[S]o far as a man has a power to think or not to think, to move or not to move, according to the preference or direction of his own mind, so far is a man free. Wherever any performance or forbearance are not equally in a man’s power,
wherever doing or not doing will not equally follow upon the preference of his mind directing it, there he is not free... (1974, II, XXI, p. 8)

David Hume (1711-1776) also sees freedom as a matching of our actions and our desires:

By liberty, then, we can only mean a power of acting or not acting according to the determinations of the will; that is, if we choose to remain at rest, we may; if we choose to move, we also may. (1955, VIII, Part I, p. 104)

It is natural to interpret Hobbes, Locke, and Hume as requiring that the will cause free actions, although, as Wegner observes (pp. 13-14), Hume emphasized that we learn the extent of our will only with great difficulty. It seems likely that conscious will is the only sort they recognized as pertinent to responsibility.

Nonetheless, it is unclear that the plausibility of compatibilism per se requires a causal relation between conscious will and our actions as opposed to a non-causal relation, especially when we give a role to conscious deliberation via distal causation. Perhaps a better requirement would be that our actions be in character with who we are, thereby demanding that morally responsible actions cohere with stable psychological dispositions. This relation might be more important to responsibility than is causation by conscious will; actions caused by aberrant, out-of-character exercises of conscious will may be poorer reflections of the nature of the subject than actions caused by steady, in-character distal causes.

This leads directly to the recent hierarchical compatibilism of Frankfurt (1971) and the reasons-sensitive compatibilism of Fischer (1994). Suppose persons’ higher order evaluations of their lower-level desires agree with their volitional behavior but do not cause the actions because of ICW. Frankfurt could say that in this case the fact that persons’ unconscious cognitions reflect who they are and what they value secures moral responsibility, even if their beliefs about the exact etiology of their actions are wrong. Or suppose that persons were responsive to good reasons in the sense that they will act in a certain way just in case they have good reasons for acting that way, despite the truth of ICW. In this case, Fischer could acknowledge ICW as an interesting phenomenon but deny that it destroys compatibilist moral responsibility. In sum, compatibilists could say that as long as persons act according to their characters, manifested in either a hierarchical (Frankfurt) or a reasons-sensitive way (Fischer), they act morally responsibly. So could libertarians such as Kane (1996) and Ekstrom (2000).

Cynics might disparage such noncausal correspondence between action and character as a lucky convergence of separate subsystems, like two trains running in parallel without anything synchronizing the engines—but notice how much this riposte looks like a traditional incompatibilist objection to compatibilism (e.g., “Your hierarchical evaluations and sensitivity to reasons themselves are determined, so you are lucky if your sophisticated cognitions correspond to an improvement in the quality of your actions”). Such an objection, though irrefutable, strikes me as antithetical to the compatibilist motivation since at least the sixteenth century, which sought to find common ground between what
compatibilists take to be the facts of science (the thesis of determinism) and moral attitudes. If ICW accurately captures what science reveals about action, undogmatic compatibilists could, with some sense of historical mission, adjust the relata of what needs to be shown compatible.

**Reply 3: Conscious Will As Epiphenomenon**

Wegner invites a third phase to this defense of moral responsibility with his speculation that conscious will may be epiphenomenal (p. 97; Ch. 9). Wegner suggests that unconscious brain states might cause both actions and thought (our awareness of willing those actions):

> The coupling of thought and action over time...is really quite remarkable if the thought is *not* causing the action, so there must be some way in which the two are in fact often connected. (p. 97)

This hypothesis makes our conscious will a reliable indicator of our actions even though it is not the cause of our present actions. This means that conscious will, although denied the glory of driving the behavioral engine, is physically necessary to it.

Consider the epiphenomenalist hypothesis that our unconscious neural states cause our actions and conscious will. According to most views of causation, if one event causes a second, then if the second had not occurred, the first would not have occurred, either (e.g., if the water flowing over the rocks in the brook causes a babbling sound, then were the babbling sound to stop, the water no longer flows). So, assuming the relation is causal, without the awareness of consciously willing (A) (the epiphenomenon), the unconscious brain might not have been such that it caused (A). To call a mental event “epiphenomenal” may be to demean it in common parlance, but common speech is slanderous on this point. Epiphenomenalism does not say that the effect might just as well have never happened.

Let me contrast two extremes by which conscious will may be related to action (excluding, as ICW does, the possibility that conscious will causes action). Wegner seems drawn toward each.

**Case 1:** Conscious and unconscious neural states cause both action (A) and directly a conscious awareness of choosing (A), which the subject mistakes as the cause of (A) [conscious will as epiphenomenon].

**Case 2:** Conscious and unconscious neural states cause action (A) but do not directly cause the conscious awareness of choosing (A). Instead, the conscious awareness of choice is idiosyncratic and mediated—caused by faulty attributions, adherence to mistaken theories of personal agency, perhaps being imputed to the virtual “I” that our brains fabricate [conscious will as attributional]. If our mistaking the epiphenomenal conscious awareness of choice to be the cause of (A)
in Case 1 is illusory, our mistaking of the attributional awareness of choice as the cause of (A) in Case 2 is delusional.

In both cases I err if I take my conscious will to cause my actions, but the difference is important. If, in Case 1, my faulty awareness is a law-like effect caused by the conscious and unconscious neural states that caused (A), then had the awareness state not occurred, neither would the conscious and unconscious neural states, the cause of (A). So, in Case 1, our awareness, epistemologically faulty though it is, is physically necessary for the actions we think we are controlling consciously. This law-like physical necessity may be enough, according to compatibilists and libertarians, to sustain moral responsibility, especially if our awareness serves as a contributing distal cause or influence for this type of action (Reply 1) and makes the action teleologically intelligible (Reply 2).

Case 2 creates difficulties for anyone who wishes to accept ICW and defend responsibility. If the relation between our conscious and unconscious neural states and our erroneous conscious awareness of choosing (A) is not epiphenomenal, then our mistaken awareness is not connected in a law-like way to our A-ing. This raises the specter that our purported awareness is fancifully attributional, inviting the outré cases Wegner documents throughout his book (dissociative identity disorder, hypnotic trances, and voodoo). In this case, conscious will would be less able to serve in the defense of moral responsibility.

Here is an example that challenges the epiphenomenalist reply. Suppose as jurors we find a defendant guilty due to some epistemically worthless subliminal visual cues (e.g., the defendant’s squinty eyes), yet attribute our verdict to our evaluation of the evidence presented at the trial. In this case we are deluded and not morally responsible for our decision. Now suppose a credible source convinces us that we are wrong about the etiology of our verdicts but offers us the epiphenomenalist account. We thereby become convinced that our conscious will (our awareness of weighing the evidence) faithfully reflects the way our conscious and unconscious neural states drove our verdict and conclude that we are responsible after all. Surely we would still be wrong in thinking we are responsible.

We should not make too much of this example, however. Here the contents of the etiology of our action (our negative evaluation of the defendant’s eyes) and our confabulated explanations (our rehearsing of the state’s evidence) diverge so widely that there is no single position that represents where we stand (as Frankfurt might say) on the question of the defendant’s guilt: not in terms of distal or proximate causation, teleological intelligibility, our stable values, our deep characters, or our sophisticated cognitions. The cause of our action is unavailable to our conscious attempt to scrutinize it, and our conscious will is not even causally necessary to it. Moral responsibility would break down—not because of ICW, but because of our fundamental irrationality (at this point we run into the part of Wegner’s thesis I have purposely avoided: the rejection of responsibility due to the breakdown of agency).
The mere logical possibility of such divergences, though, does not hurt my claim that, even if ICW is true—if conscious will is an epiphenomenon of the real unconscious causes of actions—we may be morally responsible. Moreover, we should not reject my claim simply by citing thought experiments in which the pronouncements of persons’ conscious will are deeply separated from their actual driving motivations. To argue from such cases one would have to claim that such divisions are the rule, not exceptions. This is a claim that Wegner seems drawn toward when he envisages hypnotism, spirit possession, and dissociative identity disorder, but he rejects when he discusses the evolutionary advantages that epiphenomenal conscious will provides as a source of information about our ownership of our behavior (Ch. 9).

Conclusion: The Convergence of These Replies

Wegner’s ICW challenges our acceptance of moral responsibility, given the sensible premise that we are responsible for our volitional actions only if we cause those actions by our conscious choices. Unable to evaluate ICW, I have tried to see whether we can deny the sensible premise, apply some philosophical themes, and see whether we can accept both ICW and responsibility.

One could object that my lengthy discussion is otiose because simple, commonsensical counterexamples can suffice. For example, imagine you are a bank robber who plans a robbery. You go through the details of the deed and discuss with your partner what to do if anyone tries to stop you. You decide to shoot to kill. The day of the robbery comes, the guard at the bank pulls out his gun, and you shoot him. Now, even if we accept ICW we have no reason at all to think that you are not morally responsible for the shooting.3

I agree that this sort of thought experiment seems to count against Wegner’s conclusion about responsibility. However, rather than replacing my replies, the example works only to the extent that my replies are acceptable. Even if one accepts the thought experiment, one would like to understand why it works. My explanation is that my three retorts to Wegner translate into a list of conditions that support the objection.

Assuming the robber’s pulling of the trigger was not proximately caused by conscious will, a distal ancestor with the same violent thought entered into it. So, for compatibilists, the robber’s preexisting plan to shoot to kill was a distal cause (for libertarians, a distal probabilistic cause), which made the robber responsible (Reply 1). Because the shooting was explainable by reference to the robber’s beliefs, desires, and plans, the action satisfies the teleological intelligibility condition available to compatibilists and libertarians and was an accurate indicator of the robber’s character (Reply 2). Finally, if the robber’s false awareness of consciously causing the firing of the gun was a law-like effect (epiphenomenon) of the brain processes that really caused the firing, then his murderous conscious will

---

3 I owe this example to Saul Smilansky.
was necessary to the action, even though it did not cause it (Reply 3). Thus, replies 1-3 explain the force of the thought experiment.

Suppose next that the murderer met none of the conditions cited in my replies; suppose that the act did not have the right distal background, was not teleologically intelligible and was out of character, and the conscious will was not caused by the act’s real causes but was entirely confabulated. If none of these conditions are met, then I think the robber is not morally responsible.

Finally, what happens to the thought experiment (and moral responsibility generally) if only some of the conditions given by the three replies are met? For example, what if the action has an appropriate distal cause or influence but is out-of-character and the conscious will that accompanies it is confabulated? Or what if the conscious will is an epiphenomenon of the causes of the action but the act is neither teleologically intelligible nor in character? In such cases we would not know what to say. Such possibilities make me think that Wegner’s ICW, besides being its own powerful threat to moral responsibility, reminds philosophers that their elegantly-crafted accounts stand at risk of embarrassment by rude scientific facts. Undogmatic philosophers should not assume that moral responsibility is immune from serious dangers such as those Wegner has raised.4

References


Libet, B. (1999). Do we have free will? In Libet, et. al. (Eds.), The volitional brain (pp. 47-57). Thorverton, UK: Imprint Academic.

Libet, B. (2002). Do we have free will? In R. Kane (Ed.), The Oxford handbook of free will (pp. 551-564). New York: Oxford University Press.


4 I have argued for this in Double, 1991.