Oxford Handbooks Online

Free Will and Science

Alfred R. Mele

The Oxford Handbook of Free Will: Second Edition (2 ed.)
Edited by Robert Kane

Print Publication Date: Jul 2011 Subject: Philosophy, Metaphysics, Philosophy of Science
Online Publication Date: Sep 2012 DOI: 10.1093/oxfordhb/9780195399691.003.0026

Abstract and Keywords

This article examines the work of two figures in fields whose work has had a significant impact on recent free-will debates, neuroscientist Benjamin Libet and psychologist Daniel Wegner. Libet's groundbreaking experimental studies on human subjects relating brain activities to the appearance or production of conscious experience, volition, and willed action have been much discussed by philosophers and scientists over the past few decades and have influenced subsequent scientific research on these subjects. The second half of the article deals with the arguments of psychologist Daniel Wegner, whose book, *The Illusion of Conscious Will*, has had a significant impact on free-will debates since its publication. Wegner argues that our experience of conscious control over our willed actions is an illusion. Wegner appeals in part to the Libet experiments and other neuroscientifc experiments on voluntary action.

Keywords: free-will debates, Daniel Wegner, Benjamin Libet, neuroscience, psychology, voluntary action

ONE argument for skepticism about free will features the belief—defended by Daniel Wegner (2002, 2008) and Benjamin Libet (1985, 2004), among others—that conscious intentions (and their physical correlates) never play a role in producing corresponding overt actions.[1] This chapter examines alleged scientific evidence for the truth of this belief and reviews some recent philosophical work on this alleged evidence.

Because the expression “neural correlate” is used in various distinct senses in the literature, I avoid it here. “Physical correlate” is, I hope, a relatively innocuous technical term. From a physicalist, neuroscientific point of view, proof that the physical correlates of, for example, a particular intention were among the causes of a particular action constitutes proof that the intention was among the causes of the action. It is primarily philosophers who would worry about the metaphysical intricacies of the mind-body problem despite accepting the imagined proof about physical correlates, and the relevant

argumentation would be distinctly philosophical.[2] In this chapter, I focus on empirical work at the expense of metaphysics.

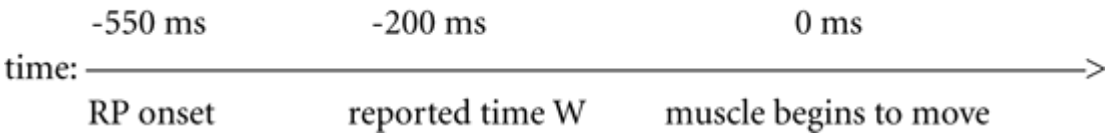
Libet's Work: Introduction

Libet (1985, 536) contends both that "the brain 'decides' to initiate or, at least, prepare to initiate [certain actions] before there is any reportable subjective awareness that such a decision has taken place"[3] and that "If the 'act now' process is initiated unconsciously, then conscious free will is not doing it" (Libet 2001, 62; also see 2004, 136). He also contends that once we become aware of these decisions, we can exercise free will in vetoing them (Libet 2004, 137–49). Libet has many critics and many supporters. Some people follow him part of the way: They accept the thesis about (p. 500) when and how decisions are made but reject the window of opportunity for free will as illusory (Wegner 2002, 55, Hallett 2007).

In some of Libet's studies, subjects are regularly encouraged to flex their right wrists whenever they wish. In subjects who do not report any "preplanning" of flexings, electrical readings from the scalp (EEGs)—averaging over at least forty flexings for each subject—show a shift in "readiness potentials" (RPs) that begins about 550 milliseconds (ms) before the time at which an electromyogram (EMG) shows relevant muscular motion to begin (Libet 1985, 529–30). These are "type II RPs" (531). Subjects who are not regularly encouraged to act spontaneously or who report some preplanning produce RPs that begin about half a second earlier—"type I RPs." The same is true of subjects instructed to flex at a "preset" time (Libet, Wright, and Gleason 1982, 325). (According to a common use of "readiness potential" [RP], it is a measure of activity in the motor cortex that precedes voluntary muscle motion and, by definition, EEGs generated in situations in which there is no muscle burst do not count as RPs. Libet's use of the term is broader. For example, because there is no muscle burst in the veto experiment described later, some scientists would not refer to what Libet calls "the 'veto' RP" [538] as an RP.)

Subjects are also instructed to "recall … the spatial clock position of a revolving spot at the time of [their] initial awareness" (Libet 1985, 529) of something (*x*) that Libet variously describes as a decision, intention, urge, wanting, will, or wish to move.[4] (The spot makes a complete revolution in under three seconds.) On average, in the case of type II RPs, "RP onset" precedes what subjects report to be the time of their initial awareness of *x* (time W) by 350 ms. Reported time W, therefore, precedes the beginning of muscle motion by about 200 ms. The results may be represented as follows:

Libet's results for type II RPs



(Libet [1985, 531, 534] finds evidence of what he regards as an error in subjects' recall of the times at which they first become aware of sensations. Correcting for it, time W is -150 ms.)

Again, in Libet's view, consciousness opens a tiny window of opportunity for free will in his subjects. If a subject becomes aware of his decision or intention at -150 ms, and if by -50 ms his condition is such that "the act goes to completion with no possibility of its being stopped by the rest of the cerebral cortex" (Libet 2004, 138), his window is open for 100 ms. Libet (1999, 54) writes: "The role of conscious free will [is] not to initiate a voluntary act, but rather to control whether the act takes place. We may view the unconscious initiatives as 'bubbling up' in the brain. The conscious-will then selects which of these initiatives may go forward to an (p. 501) action or which ones to veto and abort." His position on vetoing is discussed below ("Vetoing and Free Will").

Some Conceptual Background

A sketch of some conceptual background will facilitate an assessment of Libet's work. I start with the concept of deciding to do something (i.e., "practical deciding"). (Deciding that something is true [i.e., "propositional deciding"] is a distinct phenomenon.) Like many philosophers, I take "deciding to A" to be an action—as I see it, a momentary action of forming an intention to A (Mele 2003, ch. 9). The momentary action is, more fully, a mental action of "executive assent to a first-person plan of action" (210), in which the plan may be as simple as a prospective representation of flexing one's right wrist. Deliberating about what to do is not a momentary action, but it must be distinguished from an act of deciding that is based on deliberation.

This conception of practical deciding does not entail that all intentions are formed in acts of deciding. In fact, many intentions seem to be acquired without being so formed. For example, when Al unlocked his office door this morning, he intended to unlock it. But because he is in the habit of unlocking his door in the morning and conditions were normal, nothing called for a decision to unlock it. If Al had heard a fight in his office, he might have paused to consider whether to unlock the door or walk away, and he might have decided to unlock it. But given the routine nature of his conduct, there is no need to posit an action of intention formation in this case. His intention to unlock the door may have arisen without having been actively formed. If, as I believe, all decisions about what to do are prompted partly by uncertainty about what to do (Mele 2003, ch. 9), in situations in which there is no such uncertainty, no decisions will be made. Even so, intentions may be acquired in these situations.

Some decisions and intentions are about things to do straightaway. They are "proximal" decisions and intentions. Others—"distal" decisions and intentions—are about things to do later. Al's decision to phone Bob now is a proximal decision; his decision to phone Beth

tomorrow evening is a distal decision. The scientific work on decisions and intentions to be discussed here focuses on the proximal variety.

Deciding to do something should be distinguished from wanting (or having an urge) to do it. Sometimes people want to do things that they decide not to do. And often, when people want to do each of two incompatible things—for example, meet some friends for lunch at noon and go to class at noon—they settle matters by deciding which one to do. Just as deciding should be distinguished from wanting, so should intending. Intending to do something is more tightly connected to action than is merely wanting to do it.

(p. 502) The account of practical deciding sketched here is not the only one in the philosophical literature. (For critical discussion of alternative accounts, see Mele [2003, ch. 9].) For present purposes, a virtue of the account just sketched is that it is consonant with Libet's apparent conception of practical deciding.

Type II RPs, Conscious Decisions, and Actions

A brief description of my own experience as a subject in a Libet-style experiment will give readers a better feel for the subjects' task. I wanted to conduct myself as a naŧve subject might. My plan included waiting for something like a conscious proximal urge to flex to emerge and flexing in response to it. Of course, I also planned to attend to the clock and to report, after flexing, where I believed the revolving spot was when my conscious urge emerged. However, because conscious proximal urges to flex did not emerge in me, I altered my plan. My new plan was to say "now!" silently to myself, to flex straightaway in response to that silent speech act, and then, after flexing, to report where I thought the spot was when I said "now!" (I reported on this by moving a cursor to a point on the clock.)

I thought of the "now!" as being in the imperative mood. I thought of my conscious "now!"-sayings as conscious self-commands and as expressions of conscious proximal decisions to flex. *Must* I have been wrong about that? If I proximally decided to flex, must I have done that unconsciously? Here one must be careful not to confuse unconsciousness of *causes* of a decision with unconsciousness of the decision. As it happens, I believe that all actions—including decisions—are caused (Mele 2003). If I had been asked what caused my proximal decisions to flex, I would not have said "nothing." I might have mentioned some relevant factors: for example, my desire to have a relevant conscious event to report after each flexing and my desire to wait only a few seconds between flexings. But I certainly would have admitted that I did not experience anything as a proximal cause of any of my "now!"-sayings (or conscious proximal decisions to flex, if that is what my "now!"-sayings expressed). The point to be emphasized is that this is compatible with my having consciously proximally decided to flex. I consciously said "now!" to myself (many times); and my not being conscious of the proximal causes of

those silent speech acts does not prevent those speech acts from expressing decisions I was consciously making at the time. (For a model of conscious deciding, see Mele [2009, 40–44].)

As Daniel Dennett (2003, 228–42) observes, Libet's subjects must deal with pairs of conscious experiences in their reporting task. In my case, for example, the pairs were composed of conscious experiences of saying “now!” and conscious experiences of clock positions of the revolving spot. Dennett believes that the type II RP “is a highly reliable predictor” of flexing (229), and he contends that it is possible that Libet's sub (p. 503) jects were conscious at -550 ms of their decisions to flex and that, owing partly to their having to keep track of and compare a pair of experiences, it seemed to them as though they were first conscious of the decisions around -200 ms (234–36). (For reply to an earlier suggestion by Dennett along these lines, see Libet [2004, 59–67]. For a discussion of grounds for skepticism about the accuracy of subjects reports about when they became conscious of their proximal decisions or intentions, see Mele [2009, ch. 6].)

Even if, as Dennett (2003, 229) says, the type II RP, which begins at -550 ms, is “a highly reliable predictor” of flexing, is the brain activity registered by, say, the first 300 ms of this RP—call it “type 300 activity”—a highly reliable predictor of a flexing action or even a muscle burst? In fact, this is not known. In the experiments that yield Libet's type II RPs, it is the muscle burst that triggers a computer to make a record of the preceding brain activity. In the absence of a muscle burst, there is no record of that activity. So, for all anyone knows, there were many occasions on which type 300 activity occurred in Libet's subjects and there was no associated muscle burst. Some of his subjects reported spontaneously suppressing or aborting (i.e., vetoing, in his terminology) conscious urges to flex. As Libet (2004, 141) points out, “In the absence of the muscle's electrical signal when being activated, there was no trigger to initiate the computer's recording of any RP that may have preceded the veto.” So, for all anyone knows, type 300 activity was present before the urges were suppressed.

Notice that it is *urges* that these subjects are said to report and suppress. Might it be that type 300 activity is a potential cause of conscious urges to flex in Libet's subjects and some subjects make no decision about when to flex—unconsciously or otherwise—until after the conscious urge emerges? And might it be that prior to the emergence of the conscious urge, these subjects have no proximal intention to flex—not even an unconscious one? That our urges often are generated by processes of which we are not conscious is not surprising. And if we sometimes make effective decisions about whether or not to act on a conscious urge, so much the better for free will.

Someone who is confident that the brain events indicated by the first few milliseconds of a type II RP or by type 300 activity are reliably associated with a muscle burst at 0 ms might be attracted to Libet's claim that unconscious decisions to flex are made at -550 ms or to Dennett's (2003) suggestion that decisions are made consciously at that time.

However, in light of the point just made about how the EEG data are gathered, this confidence obviously is misplaced.

How might one get evidence about whether the onset of the type II RPs at -550 ms is correlated with (unconscious) proximal decisions to flex or instead, for example, with potential causes of proximal decisions or intentions to flex (as suggested in Mele [2009, chs. 3, 4])? An apt question to ask in this connection is how long it takes a proximal intention to flex to generate a muscle burst. If, in fact, the brain produces proximal decisions or intentions in Libet's study about 550 ms before the muscle burst, then in his subjects it takes those decisions or intentions about 550 ms to produce a muscle burst. Is this a realistic figure?

Some reaction time studies provide relevant evidence. In one study in which subjects are watching a Libet clock, the mean time between the sounding of the “go signal” and the muscle burst is 231 ms (Haggard and Magno 1999, 104). The subjects are (p. 504) instructed to respond as rapidly as possible to the “go signal” by pressing a button. If detection of the “go signal” produces a proximal intention to press the button, then the mean time between a subject's acquiring a proximal intention to press and the muscle burst is less than 231 ms. (Detecting a “go signal” takes time.) Notice how close this is to Libet's time W—his subjects’ reported time of their initial awareness of something he variously describes as an intention, urge, wanting, decision, will, or wish to move (-200 ms). Even without putting much weight on the exact number (231 ms), one can fairly observe that if proximal intentions to flex are acquired in Libet's studies, the finding just reported makes it look like a much better bet that they are acquired around time W than that they are acquired around -550 ms.

Someone might object that in reaction time studies, muscle bursts and actions are not produced by proximal intentions but by something else. It may be claimed, for example, that the combination of subjects’ “conditional intentions” to press whenever they detect the “go signal” together with their detecting it produces muscle bursts and pressings without the assistance of proximal intentions to press. But if this claim is accepted, a parallel claim about Libet's studies should be taken seriously. The parallel claim is that, in Libet's studies, muscle bursts and actions are not produced by proximal intentions but by the combination of subjects’ conditional intentions to flex whenever they detect a conscious proximal urge to flex together with their detecting such an urge. (In my case, the operative conditional intention might have been to flex whenever I said “now!”) Someone who makes this claim may hypothesize that the onset of the type II RPs at -550 ms is correlated with a potential cause of a conscious proximal urge to flex (or of a conscious “now”-saying, in my case). Libet's findings do not contradict this hypothesis.

Someone may contend that even if type 300 activity is a potential cause of a conscious proximal decision to flex and such a decision precedes a muscle burst by, say, 150 to 200 ms, that decision is not among the causes of the flexing. Some ways of defending this contention definitely fail. Roediger, Goode, and Zaromb (2008, 208) write: “Clearly conscious intention cannot cause an action if a neural event that precedes and correlates

with the action comes before conscious intention." This claim is surprising. Consider the following claim: Clearly, the burning of a fuse cannot cause an explosion of a firecracker if a lighting of a fuse that precedes and correlates with the explosion comes before the burning of the fuse. Obviously, both the lighting of the fuse and the burning of the fuse are among the causes of the explosion. Other things being equal, if the fuse had not been lit—or if the lit fuse had stopped burning early—there would have been no explosion. The surprising claim by Roediger, Goode, and Zaromb cannot undermine the hypothesis (Pacherie 2006, 162) that conscious proximal intentions to flex are part of the causal chain leading to the flexings of Libet's subjects.

Even if the claim that Libet's subjects have proximal intentions to flex before they think they do is not warranted by Libet's data, his idea that people have unconscious proximal intentions merits attention. Some psychologists view unconscious intentions as conceptually impossible (Wegner 2002, 18), and others disagree (Marcel 2003). I myself accept the existence of unconscious proximal intentions (Mele 2009; also see (p. 505) Nahmias 2005, 782). Unconscious proximal intentions may be at work when, for example, experienced drivers flip their turn indicators to signal for turns they are about to make. In a study in which subjects are instructed to flex whenever they feel like it *without also being instructed* to report after flexing on when they first became aware of a relevant intention, urge, or whatever, would they often be conscious of proximal intentions (or urges) to flex? Might unconscious proximal intentions to flex—and, more specifically, proximal intentions of which they are never conscious—be at work in producing flexings in the imagined scenario?

Imagine that the experiment just sketched is conducted and it is discovered (somehow) that the subjects were never or rarely conscious of proximal urges or intentions to flex.[5] Could one legitimately infer that, in Libet's own experiment, conscious urges, decisions, and intentions never or rarely had an effect on the flexings? No. One possibility is that some of Libet's subjects treat their initial consciousness of an urge to flex as a "go signal" (as suggested in Keller and Heckhausen 1990, 352). If they do, the conscious urge seemingly has a place in the causal process that issues in the flexing. Another possibility is that some subjects treat the conscious urge as what may be called a "decide signal"—a signal calling for them consciously to decide right then whether to flex right away or to wait a while. If that is so, and if they consciously decide to flex and execute that decision, the conscious urge again seemingly has a place in the causal process, as does the conscious decision.

Perhaps it will be suggested that even if a subject treats a conscious urge to flex as a "go" or "decide signal," that urge has no place in the causal process that issues in a flexing because "a neural event that precedes and correlates with the action comes before" the conscious urge (Roediger, Goode, and Zaromb 2008, 208). But the suggestion here suffers from the same problem as the surprising claim about "conscious intention" discussed earlier. Possibly, it will be claimed that by the time the conscious urge emerges it is too late for the subject to refrain from acting on it (something that Libet denies) and that is why the conscious urge should not be seen as part of the process at issue, even if subjects

think they are treating the urge as a “go” or “decide signal.” One way to get evidence about this (suggested in Mele 2009, 75–76) is to conduct an experiment in which subjects are instructed to flex at time *t unless* they detect a “stop signal.” (On “stop signal” experiments, see Logan 1994.) By varying the interval between the “stop signal” and the mean time of the completion of a full flex when there is no “stop signal” experimenters can try to ascertain when subjects reach the point of no return.[6]

Vetoing and Free Will

This section begins with a discussion of some of Libet's ideas about vetoing and ends with a discussion of the bearing of his data on the question whether people ever act freely. Along the way, another problem emerges for Libet's contention that (p. 506) unconscious decisions are made or unconscious intentions acquired at around -550 ms in studies that yield type II RPs.

Libet (1999, 52) discusses “the possibility that the conscious veto itself may have its origin in preceding unconscious processes, just as is the case for the development and appearance of the conscious will.” If having such an origin renders the proximal decision to flex unfree and the (decision to) veto has an origin of the same kind, its origin would seem to render it unfree. Libet contends that although “factors on which the decision to veto … is *based*” may “develop by unconscious processes that precede the veto … the *conscious decision to veto* could still be made without direct specification for that decision by the preceding unconscious processes” (53, emphasis in original). He also asserts that the “decision to veto” might not “require preceding unconscious processes.” Libet seems to be making two suggestions: first, although free decisions to veto have unconscious processes among their causes, these decisions are *not deterministically caused;* second, free decisions to veto are *not causally dependent* on “preceding unconscious processes.”

Libet (1985) mentions what he regards as two sources of evidence for veto power. The first is an experiment in which subjects are instructed to prepare to flex their fingers at a prearranged clock time but to refrain from actually flexing and “to veto the developing intention/preparation to act … about 100 to 200 ms before [that] time” (538). The second is subjects’ reports about unsolicited vetoing. Subjects encouraged to flex spontaneously (in nonveto experiments) “reported that during some of the trials a recallable conscious urge to act appeared but was ‘aborted’ or somehow suppressed before any actual movement occurred; in such cases the subject simply waited for another urge to appear, which, when consummated, constituted the actual event whose RP was recorded” (538). No record was made of brain activity associated with suppressed urges for a reason explained above.

The results of Libet's (1985) veto study suggest an interpretation of type I and type II RPs that is contrary to his own interpretation of them. As a first step toward seeing why, notice that Libet's claim that the subjects in this study veto "*intended* motor action" (38; emphasis added) is implausible (see Mele 1997, 322; 2009, 52–53). These subjects were instructed in advance *not* to flex, but to prepare to flex at the prearranged time and to "veto" this. The subjects intentionally complied with the request. They intended from the beginning not to flex at the appointed time. So what is indicated by what Libet refers to as "the 'veto' RP" before "about 150–250 ms before the preset time" (Libet 1985, 538)? Presumably, not the acquisition or presence of an *intention* to flex; for then, at some point in time, the subjects would have both an intention to flex at the prearranged time and an intention not to flex at that time. And how can a normal agent simultaneously be settled on *A*-ing at *t* and settled on not *A*-ing at *t*?[7]

A segment of "the 'veto' RP" resembles segments of type I RPs in cases in which subjects do flex, as Libet (1985, 538) observes. Given that this segment of "the 'veto' RP" is not correlated with a proximal intention to flex, perhaps the (p. 507) similar segments of type I RPs (and of type II RPs) also are not correlated with proximal intentions to flex. Even so, they might be correlated with potential causes of such intentions.

This idea is developed in Mele (2006a, 2009). The shape the idea takes there is based partly on the following possibilities about subjects in the veto experiment:

> perhaps a subject's wanting to comply with the instructions—including the instruction to prepare to flex at the appointed time—together with his recognition that the time is approaching produces an unconscious urge to flex soon, a pretty reliable causal contributor to an urge to flex soon, or the motor preparedness typically associated with such an urge. Things of these kinds are potential causal contributors to the acquisition of proximal intentions to flex. A related possibility is suggested by the observation that "the pattern of brain activity associated with imagining making a movement is very similar to the pattern of activity associated with preparing to make a movement" (Spence and Frith 1999, 27).[8] The instructions given to [the subjects in the veto experiment] would naturally elicit imagining flexing very soon, an event of a kind suitable, in the circumstances, for making a causal contribution to the emergence of a proximal urge to flex. (Mele 2009, 55)

The suggestion is that these same items—as opposed to proximal intentions to flex—are candidates for what the pertinent segments of type I RPs signify and that proximal intentions to flex emerge later, both in the case of flexings associated with type I RPs and in the case of flexings associated with type II RPs (Mele 2009, ch. 3). And again, the reaction time study discussed earlier provides independent evidence about when proximal intentions emerge that places their emergence much closer to the muscle burst than -550 ms. (For new evidence, see Trevena and Miller [2010].)

How might Libet's studies bear on free will? In instances of what has been termed “the liberty of indifference,” agents are, in Kane's (1996, 108) words, “equally attracted to more than one option.” For example, Ann may be equally attracted to the corn flakes and the wheat flakes on her breakfast menu. Arguably, she may freely choose one of the two items even though nothing important hinges on her choice. The choice of a moment to begin flexing from among an array of similar moments may be similar enough to a choice of cereal in an instance of the liberty of indifference that theorists who see the latter choice as possibly free may take the same view about the former choice. On a latitudinarian conception of free will, Libet's studies may have some bearing on free will. Of course, even if they do, their bearing may be restricted to a relatively unimpressive range of free decisions—free proximal decisions in the sphere of the liberty of indifference (see Mele 2009, ch. 4). Generalizing from results obtained in this domain to, for example, a view about distal decisions made about important issues in situations of a very different kind would be extremely bold, to say the least. Even so, Libet (1985) is inclined to generalize: “our overall findings do suggest some fundamental characteristics of the simpler acts that may be applicable to all consciously intended acts and even to responsibility and free will” (563).

(p. 508) Wegner's Work: Introduction

Daniel Wegner attempts to support his claim that conscious intentions are not among the causes of corresponding actions in two general ways. One line of argument features Libet's studies. The other, as Richard Holton (2004, 219) interprets it, “is a version of the argument from illusion.” Because I discussed Libet's work at length in previous sections, I focus on Wegner's second line of argument here. If Wegner is right about conscious intentions, then if only beings whose conscious intentions sometimes are among the causes of corresponding actions are capable of acting freely, free will is an illusion.

A variety of studies provide evidence that, in some circumstances, people are not conscious of some of their actions; in others, people believe they intentionally did things that, in fact, they did not do; and in yet others, people do things “automatically” and for no good reason. This section reviews some such findings. Assessment of their implications is reserved for subsequent sections. Some background on epiphenomenalism sets the stage.

The thesis that although all mental events are caused by physical events, no mental events are among the causes of any physical events may be termed “philosophical epiphenomenalism.” Some scientists appeal to findings of the sort to be reviewed here to support what they call “epiphenomenalism” about intentions. However, what they mean by this word in this connection is not what philosophers mean by it (see Bayne 2006, 182; Hohwy 2004, 395–96; Holton 2004, 219; Nahmias 2002, 530, 537). Attention to the difference helps forestall confusion.

Let “proximal intentions*” name a collection composed of proximal intentions, their acquisition, and their persistence. Suppose that all proximal intentions* are caused by physical events but no proximal intentions* are among the causes of any physical events. Suppose also that physical correlates of proximal intentions* sometimes are among the causes of physical events—for example, bodily motions involved in overt intentional actions. Although this pair of suppositions does not contradict philosophical epiphenomenalism, it does contradict a scientific epiphenomenalism according to which neither proximal intentions* nor their physical correlates are among the causes of bodily motions. The scientific epiphenomenalism at issue here extends to the physical correlates of proximal intentions*: The claim at issue is that neither proximal intentions* nor their physical correlates are among the causes of physical events that proximal intentions* are thought to cause—those involved in corresponding overt intentional actions.

I turn to data. Wegner (2002, 195) discusses the practice of “facilitated communication,” in which a “trained facilitator” holds the hand of “an impaired client … at a computer keyboard.” The clients are people with disorders that hamper speech, such as autism or cerebral palsy. Facilitators are supposed to help clients express themselves without influencing which keys they press or touch, and there is considerable evidence that this is what many of the facilitators intended to do and believed (p. 509) they were doing. “It was often [apparently] found that individuals who had never said a word in their lives were quickly able to communicate, typing out meaningful sentences and even lengthy reports” (196). But it was discovered that the clients’ “responses actually originate with the facilitators themselves” (197). The facilitators controlled what was typed—without realizing that.

Some actions that people do not realize they are performing are detectable with sensitive devices. Wegner (2002, 122) mentions studies done in the late nineteenth century with an automatograph, a device consisting of “a piece of plate glass resting in a wooden frame, topped by three brass balls, upon which rested another glass plate.” There also is a screen between the participant and a recording device that is attached to the automatograph. Wegner reports “some remarkable regularities” (123). “Asked to count the clicks of a metronome, … one person showed small hand movements to and fro in time with the rhythm.” Someone “asked to think of a building to his left … slowly moved his hand in that direction.” A man who was invited to hide a knife in the room and then told to think about the object moved his hand in the knife's direction “over the course of some 30 seconds.”

People suffering from a certain kind of damage to the frontal lobes display “utilization behavior” (Lhermitte 1983). An examiner touches a brain-damaged patient's hands with an empty glass and a pitcher of water or a pack of cigarettes and a lighter (Wegner 2002, 122). “The frontal-damage patients may grasp the glass and pour it full from the carafe” or light a cigarette. “One patient given three pairs of eyeglasses donned them in sequence and ended up wearing all three.” Wegner writes: “it is as though … the idea of

the act that is suggested by the object is enough to instigate the action.” The actions at least resemble automatisms.

In some experimental situations, people are caused to believe that they intentionally did things they did not in fact do. In one study (Wegner and Wheatley 1999), a confederate and a subject, both of whom are wearing headphones, jointly operate a computer mouse on which “a 12-centimeter square board” is mounted (487). About fifty tiny objects are displayed on a computer monitor, and the mouse controls the movement of a cursor over the display. Subjects are asked how much they “intended” to make a stop of the cursor on an image (488). When subjects hear the name of an image in the display (e.g., “swan”) very shortly before the cursor stops on that image, they give, on average, a higher “intended” rating to the stop than they do under other conditions, even though, in fact, the confederate is stopping the cursor on that image. (For an instructive critique of this study, see Malle 2006, 223–24.)

Studies and findings such as the ones described here are sometimes taken to support the claim that actions never have conscious proximal intentions* or their physical correlates among their causes. This is the thesis of “scientific epiphenomenalism” about conscious proximal intentions*. Now, it is true that the studies and findings indicate that people sometimes perform actions of which they are not conscious, sometimes do things for no good reason, and sometimes believe they intentionally did things they did not actually do. But how are these truths supposed to lead to scientific epiphenomenalism about conscious proximal intentions*?

(p. 510) One route that Wegner (2002, 144) maps features the proposition that all actions are caused in basically the same way. If some actions are performed in the absence of conscious intentions to perform them and all actions are caused in basically the same way, that basic way includes neither conscious intentions to perform the actions at issue nor the physical correlates of such intentions. (Only existing conscious intentions have existing physical correlates.) Why then do we even have conscious intentions? Why did we evolve in such a way as to have them? Wegner's (341) reply is that we have conscious intentions because they give us a sense of which of the things we do we are responsible for.

Whether all actions are caused in basically the same way depends on how “basically the same way” is to be read. For example, if what is meant is simply that all actions have brain events among their causes, the claim is true (in my opinion). But, of course, this leaves it open that some of the brain events that are among the causes of some actions are physical correlates of conscious intentions to perform actions of those kinds. Wegner means something much more specific—that just as people who unknowingly move a hand slowly in the direction of an object they are thinking about are caused to do so by automatic processes of which they are unaware, all actions are caused by, and only by, such processes. Wegner (2002, 97) reports that his “analysis suggests that the real causal mechanisms underlying behavior are never present in consciousness.” As usual, he has the relatively proximal causes of behavior in mind (see Nahmias 2002, 537–38). In the

following passage, Wegner goes well beyond merely *suggesting:* “it has to be one way or the other. Either the automatisms are oddities against the general backdrop of conscious behavior causation in everyday life, or we must turn everything around quite radically and begin to think that behavior that occurs *with* a sense of will is somehow the odd case, an add-on to a more basic underlying system” (144).

Conscious Will and Scientific Epiphenomenalism

As Eddy Nahmias (2002, 536) observes, Wegner's defense of his “illusion” thesis about “conscious will” is focused on *proximal* intentions. This is not surprising given Wegner's assertion that “*Intention* is normally understood as an idea of what one is going to do that appears in consciousness just before one does it” (18, emphasis in original). This assertion plainly does not apply to distal intentions. (Nor does it identify a sufficient condition for something's being an intention. As you are driving, another driver cuts you off. The following idea of what you are “going to do … appears in consciousness just before” you hit his car: “Oh no! I'm going to hit that car.” The idea expresses a prediction, not an intention; and “intention” definitely is not normally understood in such a way that this idea is an intention.) If Wegner intends his “illusion” thesis to apply even to distal intentions, he has done (p. 511) little to support that application. In this section, the spotlight remains where Wegner shines it—on proximal intentions.[9]

Some philosophers express puzzlement about what Wegner means by “will,” “conscious will,” and “the experience of conscious will,” and they float various interpretations (Bayne 2006; Holton 2004; Mele 2004). As Bayne (2006, 170) observes, the distinction between willing and the experience of willing sometimes seems to disappear in Wegner's work. Holton (2004, 220) points out that even if “conscious willings [to *A*] … contain an element that is extrinsic to the causal process” that issues in *A*-ing, the extrinsic element may be “the element that makes the willing conscious, rather than being the willing itself” (also see Mele 2004, 206, 209–10). Holton's point harks back to the discussion in “Type II RPs, Conscious Decisions, and Actions” (above) of work that may be done by unconscious proximal intentions. If, for example, conscious proximal intentions are understood, straightforwardly, as proximal intentions of which the agent is conscious, proximal intentions* (or their physical correlates) may, in some cases, do action-producing work in which consciousness of the intentions (or the physical correlate of the consciousness) is not involved. (For a modest reading of “conscious intention” designed to accommodate various things Wegner may mean by that expression, see Mele 2009, ch. 2.)

Several philosophers criticize Wegner's moving from the data he reports to the thesis that scientific epiphenomenalism about conscious intentions is true (Bayne 2006, 178; Nahmias 2002, 533; Pacherie 2006, 163; Ross 2006, 139). Elisabeth Pacherie (2006) voices a common complaint: “Some authors, including Wegner himself on occasion, seem

to think that the fact that the experience of conscious will can be nonveridical is evidence for the claim that conscious mental causation is an illusion. This inference [is not] compelling. To show that the experience of willing is not always errorless is certainly not to show that it is always in error" (163). Bayne criticizes both this inference and an alternative route to Wegner's thesis that is similar to the one highlighted in "Wegner's Work: Introduction" (above), which features the idea that "it has to be one way or the other" (Wegner 2002, 144)—either unconscious automatic processes are what produce all of our actions or "conscious will" does it all.

This stark formulation of the idea raises some of questions. Does it really have to be one way or the other? Do conscious proximal intentions* or decisions (or their physical correlates) sometimes benefit from automatic mechanisms in the causation of actions? What might count as evidence that conscious proximal intentions* or decisions (or their physical correlates) play a role in producing some actions?

Return to Libet's studies. Imagine a study of this kind in which subjects are explicitly instructed to make a *conscious decision* about when to flex a wrist and to flex in response to it. Can they comply with this instruction, literally interpreted? If they do comply, then it would seem that their conscious decisions (or their physical correlates) are among the causes of their flexing actions.

A scientific epiphenomenalist about conscious decisions, may reply that these subjects would have flexed even if they had unconsciously decided (or intended) to (p. 512) flex and, therefore, that the conscious decisions (and their physical correlates) played no causal role in producing the flexing actions. (Wegner cannot offer this reply if he is committed to the view that intentions and decisions are essentially conscious.) There is a serious problem with this reply. The reply implicitly appeals to the following principle: If *y* would have happened even if *x* had not happened, then *x* is not among the causes of *y*. And this principle is false. For example: Sally's mother drove her to school, and Sally arrived there at 8:00 A.M. What Sally's mother did was a cause of Sally's arriving at school when she did. This is true, even though, if Sally's mother had not driven her to school, Sally's father would have done so and delivered her there at the same time.

Might a scientific epiphenomenalist about conscious decisions claim that, in the imagined experiment, the subjects' conscious decisions were not among the causes of their flexing actions because the decisions themselves were caused by unconscious processes? A reader who is tempted to accept this claim has failed to absorb the moral of the firecracker analogy in an earlier section. The fact that *x* has a cause does not entail that *x* is not among the causes of *y*.

Wegner on Free Will

In Wegner's (2004) view, conscious will is intimately related to free will. He reports that his discussion of conscious will "has actually been *about* the experience of free will, examining at length when people feel it and when they do not. The special idea we have been exploring is to explain the experience of free will in terms of deterministic or mechanistic processes" (656, emphasis in original).

In a discussion of Wegner's work, Dennett (2003, 222) writes:

> If you are one of those who think that free will is only *really* free will if it springs from an immaterial soul that hovers happily in your brain, shooting arrows of decision into your motor cortex, then, given what *you* mean by free will, my view is that there is no free will at all. If, on the other hand, you think free will might be morally important without being supernatural, then my view is that free will is indeed real, but just not quite what you probably thought it was.

Dennett adds that, despite his admiration for Wegner's work, he sees Wegner as "the killjoy scientist who shows that Cupid doesn't shoot arrows and then insists on entitling his book *The Illusion of Romantic Love*" (224). One moral to take away from this is that if one sets the bar for free will (that is, for the power or ability to act freely) ridiculously high, the thesis that people sometimes act freely should strike one as ridiculous.

Wegner (2008, 234) writes:

> Experience of apparent mental causation renders the self magical because it does not draw on all the evidence. We don't have access to the myriad neural, cognitive, (p. 513) dispositional, biological, or social causes that have contributed to the action—nor do we have access to the similar array of causes that underlie the production of the thoughts we have about the action. Instead, we look at the two items our magic selves render visible to us—our conscious thought and our conscious perception of our act—and believe that these are magically connected by our will. In making this link, we take a mental leap over the demonstrable power of the unconscious to guide action … and conclude that the conscious mind is the sole player.

Obviously, even people who believe that some of their conscious intentions play a role in causing some of their behavior should not believe that "the conscious mind is the sole player." After all, among the things that play a role in causing our intentions are events in the external world. And if, for example, conscious proximal intentions* play a role in causing overt actions, causal processes of which we are not conscious link them to bodily motions.

So one should set aside the magical idea that the conscious mind or self is not itself causally influenced by anything and is a direct and complete cause of some of our actions. More realistic ideas are more worthy of attention: for example, the hypothesis that conscious intentions* or their physical correlates make a causal contribution to some behavior. Again, Wegner marshals evidence that, in some circumstances, people believe they did things that, in fact, they did not do and, in others, people believe they did not do things that they actually did. But, of course, it is a long way from these findings to the conclusion that the hypothesis just formulated is false.

Some readers who believe that our intentions to *A* sometimes make causal contributions to our *A*-ings may think that if all of our decisions and intentions have causes, then we never act *freely*. Such readers should try to explain why compatibilists and event-causal libertarians are wrong about what free action is: Theorists of these kinds regard all free actions as caused, and they regard the causes of free actions as caused. (In Mele [2006a], causal theories of free action are reviewed and two such theories are developed in some detail.) Wegner (2008, 228) asks: "Why do we experience our actions as freely willed, arising mysteriously from the self, and why too do we resist attempts to explain those actions in terms of real causal sequences, events that are going on behind the curtain of our minds?" But why think of free will in terms of a magical self? Why not side with compatibilists or event-causal libertarians?

Do we ever act freely? That depends on how free action is to be understood. If (quoting Dennett 2003, 222 again) "free will might be morally important without being supernatural," then maybe we sometimes act freely. If acting freely requires the existence of something that does not exist—a supernatural, magical self—then we never act freely. But I know of no good reason to understand free action in the latter way.

I have not offered an account of free will here. My presumed audience is primarily philosophers, and most philosophers are familiar with most of the live options about how to understand free will. In my opinion, it is fair to conclude that, (p. 514) on any reasonable conception of free will, the studies and data reviewed here leave it open both that we sometimes exhibit it and that we never do. For a discussion of imaginary experimental results that would show that no one ever acts freely, see Mele (2009, ch. 8).[10]

Notes:

(1.) Overt actions are actions that essentially involve peripheral bodily motion. Libet (1985; 1999; 2004, 137–49) maintains that once we become conscious of a decision to perform an overt action, we can exercise free will in "vetoing" it. Neither the veto nor the associated refraining from acting on the vetoed decision is an overt action.

(2.) For an excellent brief critical review of various relevant philosophical positions that highlights the metaphysical nature of the debate, see Jackson (2000).

(3.) Elsewhere, Libet (1992, 263) writes: “the brain has begun the specific preparatory processes for the voluntary act well before the subject is even aware of any wish or intention to act.”

(4.) Libet, Gleason, Wright, and Pearl (1983, 627) report that “the subject was asked to note and later report the time of appearance of his conscious *awareness of ‘wanting’ to perform* a given self-initiated movement. The experience was also described as an ‘urge’ or ‘intention’ or ‘decision’ to move, though subjects usually settled for the words ‘wanting’ or ‘urge.’ ”

(5.) At the end of the experiment, subjects can be asked how often (if ever) they were aware of proximal intentions to flex. Of course, researchers may worry about the accuracy of their reports.

(6.) Time *t* can be a designated point on a Libet clock, and brain activity can be measured backward from *t*. My guess is that in trials in which there is no stop signal and in trials in which the stop signal does not inhibit a flexing, subjects will produce something resembling a type I RP. In trials in which the stop signal inhibits the onset of EMG activity, subjects might produce EEGs that resemble what Libet calls “the ‘veto’ RP.”

(7.) Try to imagine that you intend to eat some pie now while also intending not to eat it now. What would you do? Would you reach for it with one hand and grab the reaching hand with your other hand? People who suffer from anarchic hand syndrome sometimes display behavior of this kind (see Marcel 2003, 76–81). Sean Spence and Chris Frith (1999, 24) suggest that these people “have conscious ‘intentions to act’ [that] are thwarted by … ‘intentions’ to which the patient does not experience conscious access.”

(8.) Kilner, Vargas, Duval, Blakemore, and Sirigu (2004, 1299) produce evidence that, as they put it, “the readiness potential (RP)—an electrophysiological marker of motor preparation—is present when one is observing someone else's action.”

(9.) Powerful evidence that some conscious distal intentions play a role in producing corresponding intentional actions is discussed in Mele (2009, ch. 7).

(10.) Parts of this chapter derive from Mele (2008a, 2009).

Alfred R. Mele

Philosophy, Florida State University

Alfred R. Mele is the William H. and Lucyle T. Werkmeister Professor of Philosophy at Florida State University and director of the Big Questions in Free Will Project (2010–13). He is the author of Irrationality (1987), Springs of Action (OUP, 1992), Autonomous Agents (1995), Self-Deception Unmasked (2001), Motivation and Agency (OUP, 2003), Free Will and Luck (OUP, 2006), and Effective Intentions: The Power of

Conscious Will (OUP, 2009). He also is the editor or coeditor of The Philosophy of Action (1997), Mental Causation (1993), The Oxford Handbook of Rationality (OUP, 2004), Rationality and the Good (2007), and Free Will and Consciousness: How Might They Work? (2010).