Holly Andersen
History and Philosophy of Science
1017 Cathedral of Learning
Pittsburgh, PA 15260

Two Causal Mistakes in Wegner's *Illusion of Conscious Will*[1]

Abstract: Daniel Wegner argues that our feelings of conscious will are illusory: these feelings are not causally involved in the production of action, which is rather governed by unconscious neural processes. I argue that Wegner's interpretation of neuroscientific results rests on two fallacious causal assumptions, neither of which are supported by the evidence. Each assumption involves a Cartesian disembodiment of conscious will, and it is this disembodiment that results in the appearance of causal inefficacy, rather than any interesting features of conscious will. Wegner's fallacies illustrate two take-away points to heed if making claims about the causal structure of agency.

In his 2002 book, *The Illusion of Conscious Will*, Daniel Wegner argues that numerous neuroscientific results provide evidence that our feelings of conscious will are illusory: these feelings are not causally involved in the production of action, which is rather entirely governed by unconscious neural processes. He develops an alternative explanation of feelings of will as inferences we make between occurrent thoughts and behavior that happens to match the thoughts. Wegner's argument is based primarily on two sorts of neuroscientific findings: (i) Libet's work on the timing of mental events, and (ii) the existence of separate neural pathways for the processes involved in the production of action and in the conscious feelings of will. In this paper, I will argue that Wegner misuses both these kinds of evidence, by making two fallacious causal assumptions. Each of these assumptions, when made explicit, involves a Cartesian disembodiment[2] of conscious will. By Cartesian, I mean the idea that mental processes are not physically instantiated in neural processes, that conscious experience is sufficiently separate from physical processes as to stand in causal relations to them as distinct nodes. It is this

[1] Much thanks to Sandra Mitchell, Endre Begby, and Edouard Machery for their feedback on this paper. I also benefited greatly from discussions with Rick Grush and Jim Bogen. The Biocomplexity group at the University of Pittsburgh gave feedback on numerous early versions of this argument: thanks to Thomas Cunningham, Gabriele De Anna, Christopher Heckel, Elizabeth Irvine, Catherine Stinson, and Sam Thomson.

[2] This is a somewhat nonstandard usage of the term "disembodiment," used for lack of a better term: locutions such as de-physicalization or incorporealment are awkward.

disembodiment that leads Wegner to the conclusion that the will cannot be causally efficacious. As I will demonstrate, the neuroscientific evidence, properly understood, supports no such conclusion. However, there is also a more general lesson to be drawn from this: Wegner's fallacies serve as an indicator of the kind of care that must be taken when reasoning about the causal structure of agency.

I begin by explicating the basic features of Wegner's position, and how he uses the Libet-derived ordering of events and the separation of neural pathways to support his conclusion. In section 2, I expose the causal mistake involved in Wegner's use of Libet's results; section 3 looks at the causal mistake involved in the separate pathways argument. In section 4, I argue that both of these causal mistakes lead Wegner to an unintended Cartesianism, via the misrepresentation of nodes and relationships in the causal structure of agency. My concluding remarks generalize this case to several take-away points.

I.

Two distinctions are crucial to Daniel Wegner's position. The first differentiates between what he calls the empirical will and the phenomenological will.[3] There is no question that humans *do* things: pick up coffee cups, drive cars, walk, blink. Wegner attributes these manifestations of will, observed actions, to the empirical will. We need not be conscious of performing these actions; Wegner claims that we are generally aware of only a small subset of our actions. We blink without realizing it, we walk without thinking, and we even can drive home on "autopilot." It is crucial to note how Wegner is using the term "action." Many would disagree that these examples are manifestations of will, insofar as some of them are reflexive or automatic. This is part of the point Wegner wants to make: the empirical will is what is responsible for all activities, by definition. The empirical will is operationally defined as what accounts for behavior, conscious or not. Opting for a wider use of the term than most of the philosophical literature, Wegner includes in the class of actions what is normally termed activity or mere behavior.[4]

[3] Wegner, Daniel (2002) *The Illusion of Conscious Will* Bradford Books, MIT Press, p. 27. Henceforth ICW.
[4] See, for instance, Wegner, Daniel (2004) "Precis of *The Illusion of Conscious Will*" *Behavioral and Brain Sciences* 27, p. 651. Henceforth BBS.

Rather than spend time arguing how many behaviors ought not count as actions, I will simply bracket the difference criticize his argument on different grounds.

In contrast to outputs of usually unconscious behavior generated by the empirical will, Wegner sets the phenomenological will. This is the conscious aspect of will: it includes both our intentions to act, and the "feelings of oomph" we have as we act. When we are willing to report "I did that," we are reporting feelings of phenomenological will. Of all the behavior that is attributed to the empirical will, only a small subset involves phenomenological will. While the empirical will has demonstrable behavioral results – blinking, driving – Wegner claims that the phenomenological will cannot simply be assumed to have a demonstrable effect on our behavior (BBS 652). In order to find out whether or not our conscious experience of will has any causal influence on behavior, he argues that we must first clarify what is meant by the word "will," and then look to the empirical results of science to see if this kind of will is causally influential. So far, so good.

To make this clarification, Wegner introduces the second distinction mentioned above: the word "will" can have two meanings, says Wegner, and each must be evaluated separately as to its potential impact on action. The first sense of the word is will as a feeling. This feeling is what we have as we do something: imagine wanting coffee and reaching out to pick up your cup. There is a feeling of doing, of motion being initiated and guided by the conscious subject, which accompanies the action. According to Wegner, it is this feeling of doing that marks actions as our own, rather than another person's.

The second sense of the word is will as causal force of the mind – not the experience of acting, but the causal impact this experience may (or may not) have on actions. This aspect, the will as causal force, connects what the mind has decided to do with the bodily motions needed to do it. Wegner holds that this way of construing the will's causal impact – of the mind making the body do its bidding – is inappropriate. While we cannot doubt that we have the experience of acting, Wegner cautions us against

assuming that we also experience the causal force. In a vaguely Humean vein,[5] he says we have no evidence for this causal force other than the experience – we never actually *see* the will causing our actions. "As soon as we accept the idea that the will should be understood as an experience of the person who acts, we come to realize that conscious will is not inherent in action – there are actions that have it and actions that do not." (BBS 651) We have no grounds for positing an unknown connection that mysteriously turns our mental intentions into bodily movements; what we have grounds for accepting is what we experience, and we experience will only as a feeling that accompanies action.[6]

After arguing that the feeling, not the force, is the appropriate sense of the word "will," Wegner claims it is empirically and conceptually unable to give conscious will causal efficacy. The genuine causal efficacy of agency is grounded in unconscious neural processes, not in conscious experience. As regards the feeling of doing – the "oomph," as he refers to it – he employs a quasi-Humean conceptual analysis of causation to demonstrate that a feeling cannot be a cause or a force – the two meanings of "will" cannot overlap. "Causation is an event" (ICW 13), Wegner explains, not a property. Will is a feeling, and feelings cannot be causal because they would have to have causation as a property in order to do so. Therefore, our experience of will cannot be a cause of anything. Instead, we can infer from the constant conjunction of feelings of will with actions that the will causes the actions, but this inference can never be substantiated. "In the same sense, causation cannot be a property of a person's conscious intention. You can't see your conscious intention causing an action, but can only infer this from the regular relation between intention and action." (BBS 652)

This is a quasi-Humean position, rather than straightforwardly Humean, because Wegner does hold that there is a genuine causal sequence leading up to actions. The constant conjunction of intention and action is only part of the perceived, not the real, causal sequence of actions. He draws an analogy to a magic trick: the audience is led to believe in a particular causal sequence that results in a rabbit being pulled out of the hat. It looks so simple that we think there could not be any other explanation. But in reality,

[5] Although Wegner labels his analysis Humean, it contains enough un-Humean elements to give one pause (or make Hume roll over in his grave). Because I am not criticizing his position on those grounds, I follow his own labeling of the position in this regard.
[6] See also Hume p. 105.

there is another hidden causal sequence, which is both much more complicated than the perceived one, and is genuinely causal – the perceived sequence is spurious. Similarly for agency: the perceived causal sequence is of our mental intentions causing bodily actions, but this is a mistaken inference based on the constant conjunction of felt will and action. There is a real causal sequence that leads to action, but it is hidden from the view of consciousness.

This leads Wegner to posit his Theory of Apparent Mental Causation: "people experience conscious will when they interpret their own thought as the cause of their action. This means that people experience conscious will quite independently of any actual causal connection between their thoughts and actions." (BBS 654) He posits three conditions that, when they are met, result in the experience of felt will: (i) *priority* in time of our thoughts of doing something to the occurrence of action thought about; (ii) the *consistency* of our prior thoughts with the action that occurs; and (iii) the *exclusivity* of other known causes of that action. Consider the example of picking up a coffee cup: the thought, "gee, I'd like some coffee," must occur before the actual picking up of the cup; it must be consistent in the sense that the coffee cup is the object picked up; and there must be no other obvious causes of the cup being picked up, such as another person lifting it. If these conditions are met, we experience the action of picking up the cup as consciously willed by ourselves, and as caused by our conscious willing, even though, according to Wegner, that experience is the result of a spurious inference and we are entirely unaware of the genuine causal path leading to the lifting of the cup.

While Wegner's book utilizes a large assortment of results from psychology and neuroscience, I will focus on two particular kinds of evidence he interprets in favor of the theory of apparent mental causation. The first is the findings of Benjamin Libet about how subjects experience action and decision in time. The second is the use of double dissociations between experiences of will and actual behavior or action to indicate separate processing pathways for action and conscious will. These form the primary evidentiary basis for his claim that the feeling of will is illusory.

Libet's 1985 paper, "Unconscious cerebral initiative and the role of conscious will in voluntary action," sparked a heated discussion both of the role that the will plays in

action, and of the interpretation of the results and the methodology employed to arrive at them.[7] In Libet's experiments, subjects were instructed to lift their finger "spontaneously, with no preplanning," at some random point during the course of one timed minute. The subjects were instructed to note the time at which they felt the urge or decided to lift their finger,[8] by noting the position of a rotating disc on a clock face. Readiness potentials are measured at the scalp, and are an overall measure of neuronal activity. These potentials are known to ramp up, or increase, just before action is initiated. Libet recorded the subjects' readiness potentials, and compared these to their reports of when decisions were made, with adjustments for reaction time.

What Libet found is that readiness potentials began ramping up in preparation for movement consistently about 200-500 milliseconds before subjects reported the urge to move their fingers, as if the brain was already preparing for motion before the decision to move had been made. Libet took this to mean that the conscious initiative which subjects introspected was not causally efficacious in their movement, since the movement was already initiated by neurons before the decision or urge to move was reported. Libet ultimately concluded that we do not consciously initiate action.

Wegner focuses on this negative aspect of the experiments: our decision to move lacks causal efficacy in the initiation of movement. Either the decision causes the preparation for movement, or it doesn't. If it doesn't, then the preparation for movement either causes the decision or is independent of it. In this linear causal structure, the decision could not have affected the earlier preparation.

The next kind of evidence offered by Wegner concerns the mechanisms responsible for phenomenological and empirical will. Wegner takes us on a supposed search for a localized area of the experience of conscious will, to see how it might compare to the localized areas involved in generating action. Where in the brain does the experience of conscious will arise, he asks, and is it causally involved in the production of action? It is not sufficient to know what part of the brain lights up in scanners when a subject is engaged in voluntary actions, he says, because this indicates only the areas of causal sequences underwriting action, and "this sort of evidence tells us little about where

[7] Libet (1985) in *Behavioral and Brain Sciences* is followed by extensive discussion on this point.
[8] Consciously introspecting an urge to move the finger was treated as equivalent to making the spontaneous decision to move the finger.

the *experience* of will might arise" (31). The guiding idea here is that there may be one set of neural mechanisms that allow us to perform voluntary actions, and a different set of mechanisms that leads to the experience associated with these actions. Wegner looks at ear-wiggling, phantom limbs, and brain stimulation, among other results, while investigating voluntary action, finding that "the experience of will may not be very firmly connected to the processes that produce action, in that whatever creates the concept of will may function in a way that is only loosely coupled with the mechanisms that yield agency itself" (47). Wegner speaks of looking for the will in terms of looking for a "lightbulb" that flashes in accompaniment to voluntary action. When the lightbulb flashes, subjects report actions as consciously willed. He concludes this chapter by saying that no such thing has yet been found, and is unlikely ever to be found (60). Instead, the brain "shows evidence that the motor structures underlying action are distinct from the structures that allows 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" (49). The separateness of neural systems for action and experience of will, along with their timing, are taken to be sufficient evidence of the causal impotence of consciousness: "A microanalysis of the time interval before and after action indicates that consciousness pops in and out of the picture and doesn't really seem to do anything" (59).

From these two pieces of evidence, Libet's measurements of readiness potentials and the separateness of neural pathways, Wegner establishes his positive claim about Apparent Causation. In the next two sections, I will separately examine each of these evidentiary items to see if they do in fact support his conclusion.

II.

The first major causal mistake in Wegner's reasoning is pointed out by Marc van Duijn and Sacha Bem (2005) in, "On the Alleged Illusion of Conscious Will" (VDB henceforth). They explicate the mistake and then present their own positive account of how to understand agency. I will not address their positive account, but instead expand on their criticism.

van Duijn and Bem do an excellent job of cutting straight to the heart of the problem with the way Wegner uses Libet's work. Essentially, Wegner confuses causation with constituency. Intentions don't *cause* neural processes; they *are* neural processes. This is a category mistake: "saying that neuronal activity causes conscious will is therefore very much like saying that H_2O molecules cause water." (VDB 707)

There is something dubiously Cartesian about treating mental intentions as separate from and able to stand in certain causal relations to neural processes: what might such incorporeal intentions be, if they do not involve neural processes? Wegner speaks of intentions as distinct from neural processes, and of the mind as distinct from the brain (BBS 665). This is not merely a matter of conceptual distinctness, but of separate causal entities. Wegner's picture is one where intentions either cause the neural processes that lead to action, or else are causally disconnected from these processes. In light of Libet's results, Wegner thinks the latter is the only scientifically respectable answer. However, as van Duijn and Bem show, once we remind ourselves that intentions are physically instantiated, that they are constituted by neural processes, then Libet's results are innocuous. It is not possible to discriminate with readiness potentials between neuronal activity that constitutes decision making, of which the subject's report is the conclusion, and the neuronal activity that constitutes preparation for movement. Wegner committed a mistake by expecting to find a causal relationship in a relationship of constituency.

Although van Duijn and Bem don't emphasize this point, their criticism highlights a fundamentally Cartesian assumption underlying Wegner's use of Libet's results. While Cartesianism is not expressly endorsed by Wegner, in this particular case it arises out of an oversimplification of causal structure. The way Wegner parses causal variables and asks how they stand to one another renders the two variables distinct from one another. Conceiving of intentions and neural processes as distinct entities that could stand in causal relations to each other requires them not to overlap, as they would with constituency, and is the result of using a linear causal model: either the intentions come first, or the neural processes come first. By failing to capture the complexity of the constituency involved – that intentions are instantiated in the brain precisely as neuronal activity – Wegner lacks the means to model a hierarchical causal structure and ends up

with a conclusion, the causal inefficacy of conscious will, that is an artifact of the representation of results within an oversimplified causal structure.

III.

While van Duijn and Bem accurately diagnose the problem with Wegner's use of Libet's experiments, their criticism does not address his point about the separateness of neural pathways for experience of will and for action. The constituency-not-causality criticism doesn't apply to a dissociation between separate pathways: there could not be such an identity relation between them because the entire point is that they are different. Instead, I will show, by looking at other instances of separate pathways in neuroscience, that such separation need not lead to the conclusion that one pathway lacks causal efficacy, or fails to causally influence the other pathway. The evidence simply does not license such an inference. We need not be forced to either attribute complete causal control to conscious experiences of will, or else write it off as an illusion. This dichotomy, necessary to reach Wegner's conclusions, is based on another oversimplified causal assumption: that what conscious will is supposed to affect is the other pathway, the unconscious neural processes involved in action. This again involves an implicit treatment of conscious will as lacking physical instantiation in neural processes, what I've labeled his untended Cartesianism.

An excellent counterexample is the pathways involved in object recognition and spatial recognition, both part of the visual system. Separate cortical pathways for object vision and spatial vision were demonstrated by Mishkin, Underleider, and Macko (1983, MUM henceforth). Visual perception has at least two cortical pathways: the ventral stream for recognizing objects; and the dorsal stream for recognizing spatial relationships and locations (MUM 414). Lesions in the dorsal and ventral area in monkeys and humans indicate double dissociation between the two streams. Object agnosia, for instance, results from a lesion to the ventral stream: patients have intact visual sensory abilities, and are able to individuate objects, but cannot recognize what they are, nor use them as a guide in behavior (such as finding food using a previously-seen object as a guide; see Banich 2004, p. 225). Lesions in the dorsal stream do not affect object

recognition, but lead to disorders such as topographic disorientation. Patients are unable to find routes around in their environment, and may, for instance, be unable to find thei way back to their hospital room. They recognize objects, but lack the ability to use spatial location cues (Banich, 242). This separation of neural pathways, their task specializations, and their mutual contribution to vision and visually-informed action, are well-established.

The dorsal and ventral streams of vision are by no means the only such examples, as separate pathways are ubiquitous in the brain. Most importantly, such examples are not taken to indicate that either one pathway is illusory or causally ineffective, or that these pathways do not causally interact in the production of behavior. The range of interesting questions brought up by the dorsal and ventral streams of vision include the following: at what point do these streams separate?; where and how are they reintegrated to yield a unified sense of vision?; what kinds of tasks rely primarily on one stream rather than another?; and what happens when conflicting information is available to the different streams? Wegner does not ask any of these questions about agency. He concludes it must be epiphenomenal.

A primary reason for Wegner's case that conscious will is not causally efficacious in agency is the separation of pathways generating action and generating the feeling of will. Precisely the same kind of separation occurs in the case of dorsal and ventral streams of vision. This provides no reason to think that object recognition is either illusory, or causally unconnected to spatial recognition. What it does mean is that there must be a more complex causal relationship between the two than identity. Similarly, the evidence does not support Wegner's claim that experience of will, because it utilizes different pathways or neural structures than action initiation, must be either illusory or causally unconnected to action generation. The evidence radically underdetermines this conclusion. That the experience of will and the generation of behavior are not identical should not be news to anyone.

Hiding beneath Wegner's emphasis on the separateness of pathways is an assumption that if conscious will were to be causally efficacious, it would have to be causally efficacious on the other pathway, on the separate and unconscious structures

involved in producing actions.[9] This, I argue, is a fundamental misunderstanding of what conscious will is supposed to accomplish. Conscious will is not supposed to causally influence the other unconscious causes involved in action production. The component effect of conscious will on action is simply not directed at the other unconscious components of action. Rather, the component effects of conscious will are directed outward at the world; the unconscious components are also directed outwards, not intra-system to the other components. Just as the dorsal stream's causal role is not to determine how the ventral stream does its work, the causal role of conscious will should not be understood as determining how another pathway for behavior generation works. The two work in conjunction to produce a combined result; conscious will is not supposed to work by consciously affecting neural activity.

IV.

I conclude with a more general overview of the dualistic causal assumptions needed to conclude that. Let us re-examine a point from section 1. Wegner's argument has the following structure: the will should be understood as a feeling; causes are events, not properties (of feelings); and therefore feelings (of will) cannot be causes. This may be a clever way of formulating the problem, but the analysis invokes an unnecessary and unwarranted Cartesian assumption, the implicit premise that feelings cannot be events. This premise is true only if feelings are not physically instantiated. Neural events can be causes. Only if feelings are completely incorporeal, having no neural activity associated with them in any fashion, would it be acceptable to say that feelings cannot be events, and so cannot serve as causes. Wegner's conceptual analysis attempts to demonstrate a breach between experience and behavior by assuming the nonphysicality of experience. It is no surprise that, with such an assumption, one will have difficulties relating experience back to the physical actions of the body – this is the quintessentially Cartesian dilemma, and a large part of why one wants to avoid Cartesianism in the first place. The conclusion only follows, though, if we allow the implicit disembodiment of conscious

[9] In Andersen (MS), this is labeled the Micromanagement Model of conscious agency: that if conscious features of experience are to be causally efficacious, they must be directed at controlling neural processes. I mention here a criticism of Micromanagement which is expanded on in (MS).

features of agency. If we disallow this treatment of anything conscious as incorporeal, the conclusion of epiphenomenality cannot be drawn from the evidence.

This leads to take-away point 1: Node Choice Matters. The way in which causal variables are individuated, to be subsequently investigated as to their causal relations, has physical significance. Which nodes are chosen will determine to a largely unappreciated extent the results one finds.

This disembodiment is also at work in Wegner's use of Libet's results. No one disputes that there is a ramp-up of neuronal activity just prior to the conscious decision being reported. However, this would support the conclusion of the inefficacy of the decision only when conjoined with an assumption that none of the observed increase in neuronal activity is due to decisionmaking. One would have to assume that the decisionmaking, which led to the reported decision to lift a finger, was neurally unobservable, and that none of the observed increase in neuronal activity was involved in any conscious experience. These two quite dubious assumptions exemplify the physical significance of node choice, in take-away point 1. Treating the conscious decision as separate from, and standing in causal relations to, the neuronal activity surrounding it in time is to disembody the decision.

Finally, let us take another look at the claim that the existence of separate neural pathways must render one pathway causally ineffective. Wegner supposes that if conscious will is to have causal efficacy, it must causally affect some other pathway directly, specifically, the unconscious pathway involved in action production. But the causal power of agency belongs to the entire system, which includes *both* conscious and unconscious aspects; and the causal efficacy of the system is directed out towards the world, toward tasks like lifting coffeecups. The causal relationships between the dorsal and ventral streams of vision do not themselves constitute vision. Similarly, the causal relationships between the neural processes of conscious will and neural mechanisms of behavior generation do not constitute will. This is take-home point 2: Point your Causal Arrows in a Safe Direction. In this context, that means not at each other.

Take-away point 2 also applies to using an oversimplified causal representation. A linear model of causation, where there is a single chain of causes in a row, will fail to capture the intriguing causal relations seen with separate pathways (not to mention

constituency). A well-established phenomenon like the dorsal and ventral streams of vision cannot be effectively modeled in this way. There is no reason to think pathways involved in will and action would fare any better under this treatment.

It is reasonable to assume that Wegner did not intend to make Cartesian assumptions about his subject matter. His book has the tone of "going back to the science" to fix the unjustified assumptions of philosophy. So how did he end up with such a Cartesian position? His problems are artifacts of the causal representations he used, of putting a multi-level, complex causal system into a linear, single-level representation. The simplifying assumptions required to fit the extraordinarily complex nexus of conscious and unconscious neural processes into such a model ended up severing the intimate connections between action and experience. An appearance of causal inefficacy is a function of a poor representation of the causal relationships involved, not a genuine feature of conscious will. The treatment of conscious will and neural processes as distinct causal entities already carries with it a Cartesian disembodiment of the conscious will. I take it for granted that Cartesian disembodiment is *not* a desiderata of a good account of agency, scientific or philosophical. In order to avoid such disembodiment, due care must be taken to preserve structural complexities in causal representation per take-away points 1 and 2.

References

Andersen, Holly (dissertation manuscript) *The Causal Structure of Agency*.

Banich, Marie (2004) *Cognitive Neuroscience and Neuropsychology 2nd edition* Houghton Mifflin Co: New York.

Hume, David (2001) *An Enquiry Concerning Human Understanding* ed. Anthony Flew, Open Court Press: Chicago.

Libet, Benjamin (1985) "Unconscious cerebral initiative and the role of conscious will in voluntary action" *Behavioral and Brain Sciences* **8**, 529-566.

Ungerlieder, Mishkin, and Macko (1983) "Object vision and spatial vision: two cortical pathways" *TINS* October, 414-417.

van Duijn, Marc and Bem, Sacha (2005) "On the Alleged Illusion of Conscious Will" *Philosophical Psychology* Vol. 18, No. 6, 699-714.

Wegner, Daniel (2002) *The Illusion of Conscious Will* MIT Press.

-- (2004) "Precis of *The Illusion of Conscious Will*" *Behavioral and Brain Sciences* **27**, 649-692.

Wegner, Daniel (2005) "Who Is the Controller of Controlled Processes?" in *The New Unconscious* ed. Hassin, Uleman, and Bargh (2005) Oxford University Press.