

The Role of Consciousness in Free Action

PHILIP WOODWARD

1 Introduction

It is easy to prompt the intuition that free will and consciousness have something importantly to do with one another. Consider cases such as the following:

Case 1: Your spouse accuses you of having broken your commitment to the family diet by stealing cake from the refrigerator in the middle of the night, and he is in possession of damning videographic evidence. In your defense, you point to your long history of sleep-walking, and you suggest that you make an appointment with a sleep doctor.

Case 2: On a weekend, you get behind the wheel of your car and head to the grocery store. As you drive, your mind drifts, until you suddenly realize that you are *not* on your way to the store after all, but have instead followed the route you take to work each weekday.

In neither of these cases does it seem that your action (eating a piece of cake, driving to work) is the product of your free will, and this fact can be explained, in both cases, in terms of your lack of conscious awareness of what you were doing. But it is not always so intuitively clear what type or degree of conscious awareness is required for an action to count as free, as exemplified in the following cases:

Case 3: During a game of basketball, you steal the ball and make a fast break down the court. As you set up to make a shot, you spot an open teammate out of the corner of your eye, and without thinking dish the ball to her so she can complete an unimpeded lay-up.

Case 4: As you watch an advertisement on television for a charity organization, you resolve to make a generous donation, and this in contrast to your typical stinginess. Later you learn that the advertisers flashed the word “GIVE” for a duration too brief for you to have consciously processed.

In Cases 3 and 4, your actions (passing a basketball, making a donation) are significantly shaped by non-conscious processes of which you were unaware at the time of acting. As a consequence, should we say that they were not done freely?

This question is difficult to answer in the absence of a general theory of the relationship between consciousness and free will. And no consensus currently exists regarding this relationship. The reason is that the cognitive function of consciousness, on the one hand, and the necessary conditions for freedom, on the other, are matters of considerable controversy. Regarding the cognitive function of consciousness, at least the following five options have contemporary adherents (roughly in descending order of how psychologically important consciousness is):

The Selfhood Role: A subject, properly speaking, is identical with a *conscious* subject; non-conscious states and processes play only a supporting role. (While held paradigmatically by robust realists about the self, this view is by no means restricted to them. It can also be held by those who identify the self with the stream of consciousness, or with a narrative constructed out of conscious experiences.)

The Self-awareness Role: Consciousness alone affords subjects with privileged access to their own mental states.¹

The Content-Grounding Role: All of a subject's contentful mental states are individuated via their relations to actual or potential conscious states; non-conscious states count as contentful (that is, being about something, representing something) – or even as *mental* – only derivatively.²

The Integration Role: Consciousness coordinates the functioning of otherwise self-contained cognitive modules, by integrating the information processed by these modules into a central, domain-general stream of information.³ (“Domain-general” contrasts with “domain-specific,” i.e., pertaining to one type of information – visual stimuli, for example).

The Grab-Bag Role: There is a set of loosely related cognitive processes such that consciousness is always or is typically involved in their implementation. Candidates include forming a self-concept; learning new, complex behaviors, building associations between temporally separated events; performing multi-step, goal-directed activities.⁴

Regarding the necessary conditions for freedom, the following four are the most commonly discussed:

The Control Condition: An action is free only if it is generated by the agent, rather than coerced or compelled by forces external to the agent.

The Reasons-Responsiveness Condition: An action is free only if it is done for a reason, or else such that if the agent had had a different or better reason, the agent would have acted according to that reason.⁵

The Identification Condition: An action is free only if it expresses the agent's “true self” (i.e. her established commitments/values) or is such that the agent can or would *identify* with the action.

The Open-Future Condition: An action is free only if temporally prior states of the agent and her environment, in conjunction with the laws of nature, do not determine that she performs it.⁶

(I mean these formulations only as initial, rough approximations. In what follows I will note when the differences between rival formulations become relevant.)

It's possible to combine items from each of these two lists to form a wide range of general theories about the relationship between consciousness and free will. In the next section, I discuss four (families of) options: the constitution view, the causal-dependence view, the counterfactual-dependence view, and the independence view. In the final section, I review some of the most influential recent findings in the empirical sciences that touch on the

relationship between consciousness and the will, and I discuss the potential philosophical significance of these findings.

Two caveats. First, some of the literature I draw on focuses on the relationship between consciousness and moral responsibility rather than freedom. Second, while my focus will be on *phenomenal* consciousness (mental states' subjective feel), some of the literature I draw on focuses on *access* consciousness (mental states' availability for participation in various online processes). The concept of moral responsibility is not the same concept as that of freedom, nor is the concept of phenomenal consciousness the same as the concept of access consciousness. Nevertheless, the extensions of the concepts in each pair overlap enough to serve my purposes.⁷

2 Conceptual and Metaphysical Considerations

In the present section, I discuss four ways of understanding the metaphysical or conceptual connection between consciousness and free action. The first three posit a modally strong connection of some sort, but differ in how direct that connection is. I will call these three views "the constitution view," the causal-dependence view," and the "counterfactual-dependence view." The fourth view, "the independence view," is the denial that any such modally strong connection exists.

1. *The Constitution View.* According to constitution view, necessarily, all free action is constituted by consciousness. That is, free actions metaphysically consist, at least in part, in phenomenally conscious episodes of a special sort. We can get the flavor of a view of this sort by putting together two observations about action – or rather, controversial gloss on two uncontroversial observations about action.

The first uncontroversial observation is that there is, at least sometimes, something that it is like to act – a phenomenology of action.⁸ More controversial is the claim that the phenomenology of action consists in an irreducibly "executive" type of phenomenology or as Ginet (1990) puts it, a distinctively "actish" phenomenal quality that cannot be reduced to other types of phenomenology, for example the conscious representation of oneself as acting, or to a consciously represented intention to act followed by feeling one's body move.⁹

The second uncontroversial observation is that, among the many correct descriptions of a token action (flip the switch, turn on the light, alert the intruder, etc.), some do a better job than others at capturing the agent's perspective on the action, and one description in particular will capture the most basic action that the agent is trying to perform.¹⁰ More controversial is the claim that basic actions can occur in cases of total action-failure: for example, a unknowing paralytic can *try* to raise her arm – and thus really *do* something – without producing the slightest muscular movement.¹¹ If we suppose that the concept of executive phenomenology and the concept of a basic action converge on the same psychological phenomenon, what emerges is the existence of what Brian O'Shaughnessy (1973, 2003) calls "tryings": conscious experiences that consist "in doing, intentionally and with just that purpose, whatever one takes to be needed if, the rest of the world suitably cooperating, one is to perform that action."¹²

Armed with this category of executive phenomenology, we can now understand what the constitution view comes to. According to the constitution view, free actions are constituted (at least in part) by conscious episodes of an irreducibly executive type. No action that does not partly consist in an agent's consciously trying to perform it counts as free.

Why think free action is so tightly connected to consciousness? There are, I think, two types of argument that capture the appeal of the constitution view. The first argument links

together the initial items on each of the two lists I presented above (proposed cognitive roles of consciousness, proposed necessary conditions on free action). Suppose that consciousness plays the *selfhood role*: a subject, properly speaking, is identical with a *conscious* subject. To this claim add the Control condition on freedom (an action is free only if it is generated by the agent, rather than coerced or compelled by forces external to the agent). Assuming that an *agent* simply is a self or subject who is (or is capable of) acting, then it follows from these two commitments that acts are free only if generated by a conscious subject. There is more than one way to understand what it is for a conscious subject, as such, to “generate” an act, but perhaps the simplest picture is one on which the act is a part of the subject’s stream of consciousness, which is what the constitution view maintains. (In a slogan: unless you’re directly conscious of what you’re doing, you’re not in control of it.) Call this the “Selfhood Argument” for the constitution view.¹³

The second argument links together the second items on each of the two lists above. Suppose that consciousness plays the *self-awareness role*: that is, consciousness alone affords subjects with privileged access to their own mental states. To this claim add the Reasons-Responsiveness condition on freedom (an action is free only if it is done for a reason, or else such that if the agent had had a different or better reason, the agent would have acted according to that reason.) These two commitments motivate the constitution view, given the additional claim that an agent (a) acts for some reason only if she has privileged access to the reason for which she acts, and (b) has such privileged access simply in virtue of performing the action. Though there may be more than one way for an agent to have privileged access to her reasons for acting “in virtue of” performing an act, perhaps the simplest way for this to occur is if executive phenomenology itself encodes the reason for which an action is consciously executed. That is, one is only rationally guiding one’s action if, simply in virtue of consciously acting, one is aware of one’s motivating reason for acting. (In a slogan: unless you’re directly conscious of why you’re doing what you’re doing, you’re not doing it for a reason.) Call this the “Self-Awareness Argument” for the constitution view.¹⁴

2. *The Causal-Dependence View*. According to this view, necessarily, all free action is causally dependent on consciousness. The causal-dependence view shares with the constitution view the commitment that consciousness plays a part in all free actions. But it is weaker than the constitution view, because it does not require that conscious events themselves count as actions. Rather, it is consistent with the view that free actions are cause/effect pairs, and only the causes need be conscious. For example, one version of this view would be that free actions consist in physical or mental changes caused by freely formed conscious intentions. If for whatever reason a conscious intention fails to effect any change, no action has been performed (either because utterly unsuccessful tryings are impossible, or because they are possible but are not in themselves actions).

The causal-dependence view can be motivated by slightly weaker versions of the Selfhood Argument and the Self-Awareness Argument. According to the weaker version of the Selfhood Argument, all that is required for a conscious subject, as such, to generate an action is for the triggering cause of that action to be part of her stream of consciousness (the action itself need not be). And according to the weaker version of the Self-Awareness Argument, all that is required for an agent to have privileged access to her reasons for acting is for the intention that governs her action to be part of her stream of consciousness (the action itself need not be).

The Open-Future condition provides a third motivation for the causal-dependence view. According to the Open-Future condition, recall, an action is free only if temporally prior states of the agent and her environment, in conjunction with the laws of nature, do not

determine that she performs it. Many philosophers have supposed that the Open-Future condition is in prima facie tension with the Control condition: if an agent's actions are not determined by temporally prior states of hers, then how is she in control of her actions? Libertarians have responded to this challenge in a variety of ways.¹⁵ One such response that appeals directly to consciousness is the so-called "wave-function collapse" theory of mental causation, developed by Henry Stapp (2005) and Hans Halvorsen (2011).¹⁶ Physical systems can inhabit "superpositions": they can lack determinate values for some of their quantities. But, plausibly, it's not possible for conscious states to exhibit that sort of indeterminacy. According to the wave-function collapse theory, physical systems come to inhabit determinate states when they are suitably causally related to determinate, conscious states. Thus, contemporary physics may leave room for consciousness to play a fundamental role in setting the parameters for the evolution of an otherwise indeterminate, open-ended system. Advocates of the wave-function collapse theory provide support for their view by arguing that it solves the thorny "measurement problem" in quantum mechanics. Nevertheless, it remains a minority view.

There are reasons to think that both the constitution view and the causal-dependence view are too restrictive. Suppose an agent works hard to overcome a compulsion to look at his cell phone while driving. His inclination is to reach for his phone when it vibrates, but he conscientiously develops an alternative habit – giving the steering wheel a strong squeeze, say. Eventually he reacts in this way even when his attention is otherwise occupied. Is his action unfree? Certainly his action meets the Identification condition: his behavior is the manifestation of his established commitments and values.¹⁷

Consider another example. Most drivers are not aware of signaling when turning; they simply intend to turn, and signal that they are about to turn as a matter of habit.¹⁸ At the same time, their signaling is not a haphazard or reflexive action, but is done for a reason, namely, in order to alert other drivers to an immanent directional change. Thus when drivers signal, they satisfy the Reasons-Responsiveness condition. (And, of course, if they fail to signal, they can be held legally and morally responsible.)

There are two ways for the advocate of the causal-dependence view to accommodate these cases. First, she could maintain that, while the Identification and Reasons-Responsiveness conditions are necessary conditions on freedom, they are not sufficient; conscious causation of action is required as well. After all, the purpose of cultivating habits is to establish a new default way of behaving. And in an important sense, default behaviors are not freely chosen.

A second way of responding is to grant that the cases just described can be instances of free actions, so long as conscious control has played a role in the cultivation of the relevant habits. That is, the habits in question were acquired by the agent only as he has engaged in repeated, consciously controlled action. So even if his current action (squeezing the wheel, signaling a turn) is not consciously caused, it is the causal descendent of similar actions that were consciously caused. To accommodate this response, we can distinguish between two versions of the Control condition. According to the first, the exercise of control has to be *direct* (i.e., proximal):

Direct Control Condition: An action is free only if it is directly generated by the agent.

According to the second, control can be exercised in an indirect way, via the direct cultivation of a habit or the forming of an intention to carry out a long-term project:

Indirect Control Condition: An action is free only if it is directly generated by the agent or is the manifestation of a habit or standing intention that was directly generated by the agent.

These two versions of the Control condition can be used to motivate different versions of the causal-dependence view. On the strong causal-dependence view, an action is free only if it is caused by a conscious intention on the part of the agent. On the weak causal-dependence view, an action is free only if it is caused by a conscious intention on the part of the agent, or it is the manifestation of a habit that has been cultivated as the causal consequence of conscious intentions on the part of the agent.¹⁹

3. *The Counterfactual-Dependence View.* According to this view, necessarily, all free action is counterfactually dependent on consciousness. Returning to the two cases of habitual action just discussed: there is another way to understand how these cases satisfy the Control condition, a way that doesn't appeal to the agent's history. Both the direct and indirect versions of the Control condition are *actualist* versions, but the Control condition can also be understood counterfactually:

Counterfactual Control Condition: An action is free only if the agent would have generated the action had automatic mechanisms been insufficient for its occurrence.

If the Control condition is understood in this way, then it generates an even weaker version of the Selfhood Argument. Supposing, again, that consciousness plays the selfhood role (such that only conscious action-generation counts as genuine control), it follows from the Counterfactual Control condition that an action is free only if it *could have been* consciously generated.

In the cases above, non-conscious wheel-squeezing and non-conscious signal-turning both satisfy this condition. That is, perhaps what makes these actions free (if they are) is not so much the agent's history but the agent's possessing, at the time of his acting, certain capacities that may or may not be operative. We can contrast such free-but-not-consciously-caused behaviors with compulsions (e.g., tics, behaviors following hypnotic suggestion, addictive behaviors) and with unwilled bodily movements (e.g., heart-beats, saccading of the eyes), neither of which can be controlled consciously.

Likewise, both variants of the Self-Awareness Argument we have considered so far have rested on an actualist construal of the Reasons-Responsiveness condition (just as the Selfhood Argument for the constitution view and the causal-dependence view rested on an actualist construal of the Control condition). But there is also a counterfactualist version of the Reasons-Responsiveness condition.

Counterfactual Reasons-Responsiveness Condition: An action is free only if, had the agent possessed a reason to act otherwise, the agent would have acted for that reason.²⁰

So understood, the Reasons-Responsiveness condition generates a weaker version of the Self-Awareness Argument. Suppose that consciousness plays the self-awareness role, and that an agent (a) acts for some reason only if she has privileged access to the reason for which she acts, and (b) has such privileged access simply in virtue of performing the action. It will follow from the Reasons-Responsiveness condition only that it must be *possible* for an agent to consciously entertain reasons for acting otherwise than she (freely) does, not that her actual reason for acting must be conscious at the time of her (freely) acting.

The Selfhood and Self-Awareness Arguments appeal to the first two proposed cognitive roles for consciousness (respectively). The counterfactual dependence view can be motivated independently via the third: namely, the content-grounding role. John Searle has famously argued that, necessarily, all mental states with intentional content are either actually or potentially conscious, a claim he calls "The Connection Principle." A "potentially conscious"

mental state just is a capacity to bring about a *conscious* intentional state.²¹ Or, as Declan Smithies (2012) puts the idea, all intentional states are either conscious states or they are individuated by their relations to conscious states. For example, on this view, what it is for a subject to believe that Cairo is the capital of Egypt is for a subject to be capable of consciously judging that Cairo is the capital of Egypt under certain circumstances. Though a belief isn't a conscious state, it is individuated via its relation to conscious states.

Searle and his ilk argue for the Connection Principle in several ways. First, they claim that on reflection, the idea of intentional content that that is not presented to a subject makes no sense.²² Second, they argue that all attempts to explain how the content of mental states is fixed, save for those that appeal to consciousness, leave content hopelessly indeterminate.²³ Third, they argue that all attempts to explain why some systems have intentional properties while other systems do not, save for those that appeal to consciousness, imply that unthinkable things such as tornadoes and toilets have intentional states.²⁴

If these arguments are successful, then whenever an agent is motivated by a reason to do something – i.e., if her mental states encode that reason in any way – then she must be capable of consciously entertaining the content of that reason. Thus, on any version of the Reasons-Responsiveness condition, it follows from the Connection Principle that an agent acts freely only when she is capable of consciously entertaining the content of her reason for acting. Now, whatever the merits of Searle's arguments, they have failed to move the majority of philosophers of mind, who maintain on the contrary that both folk and empirical psychology rightly invoke non-conscious intentional states that have no constitutive connection to consciousness.

Finally, as King and Carruthers (2012) point out, the Identification condition on its own motivates the counterfactual-dependence view. The Identification condition is substantive only if there is a way to distinguish between an agent's established commitments and values, on the one hand, and those attitudes of hers she does not genuinely identify with, on the other (compulsive or deeply selfish desires, say). Unless the agent *would consciously endorse* some attitude of hers, it doesn't seem right to treat that attitude as part of her established commitments and values.^{25,26}

The counterfactual-dependence view grants the metaphysical possibility of an agent whose conscious states are never causally responsible for his actions, but who nevertheless acts freely. We can draw an analogy with the relationship between a project-manager and the team of employees she oversees. We can imagine that the team is so effective that at no point does the manager ever exercise any top-down control, not even at the project's outset. (The employees are so deeply committed to the mission of the corporation that they can anticipate which projects need to be implemented before being told.) Nevertheless, the manager is poised to intervene *were* the team to get off-track. If we find it plausible to say in such a scenario that the manager controls the project and that the team is responsive to the manager's goals, then we are not likely to be troubled by the fanciful epiphenomenalist possibility that the counterfactual-dependence view permits.

4. *The Independence View.* The counterfactual-dependence view is consistent with life-long epiphenomenalism on the part of an agent's conscious states, so long as the agent has the *capacity* to be conscious in certain ways. But not all theorists about free will and consciousness are committed even to a counterfactual dependence relation between free action and consciousness. There are philosophical and scientific theories of agency, on the one hand, and of consciousness, on the other, that are not obviously mutually implicating in any way; any connections that actually obtain between them are contingent.

Consider, for example, biologically based theories of agency. According to these theories, what it is to be an agent is to be an "autonomous" or self-organizing system, i.e., a system that

maintains its homeostasis over and against, but also drawing upon, environmental contingencies. The more “flexibility” a system exhibits – that is, the less its behavior is dictated by its immediate environment – the more free it is. This distinctly biological notion of agentic control requires no contribution from the system’s stream of consciousness (if it even has one).²⁷

Certain computationally based theories of agentic control are likewise silent on the role of consciousness. According to one leading computational model, control consists in the successful deployment of “comparator” mechanisms. These mechanisms exploit a feedback loop linking (a) an initial action plan, (b) the predicted outcome of that action plan, and (c) the perceived outcome of that action plan.²⁸ On this model, an agent is in control of her action when these feedback loops successfully enable her to complete her action plan, updating it in real time if necessary – picking up one’s cup of coffee without spilling it, for example. In other words, an agent controls her action to the degree that her intentions are successfully implemented, and there is no requirement that the agent be aware either of her intentions or the means by which they are executed.

Now, there are two procedures implicated by model of control that might require the assistance of consciousness. First, consciousness might be required as part of the feedback loop. For example, it might be that conscious *attention* is what enables a subject to acquire the requisite perceptual feedback.²⁹ But it’s not clear why the model would require that such attention be *conscious*. For example, when a somnambulist successfully extracts a piece of cake from the refrigerator, she seems to be satisfying all of the model’s requirements for exercising control, yet her attention to the cake remains unconscious.³⁰

The second potential role for consciousness within the comparator model of control is in fixing those intentions against which successful control is measured. That is, perhaps controlled, reason-responsive, and/or true-self-expressive action only occurs when the agent’s intention was formed via conscious deliberation (which would take us back to the causal-dependence view), or at any rate such that the agent could have consciously entertained her intention (which would take us back to the counterfactual-dependence view).

As far I can tell, however, nothing in the computational model of control itself suggests these restrictions. Now, computationalists often do make a distinction between “person-level” representations and “sub-personal-level” representations. The former are those representations that play a role in explanations of the behavior of *the person*. The latter are those representations that play a role in explanations of the behavior of *cognitive mechanisms* of the person.³¹ Is there any reason to think that the person-level/sub-personal-level distinction maps onto the consciousness/unconsciousness distinction? Many philosophers and psychologists think so. There is widespread consensus that consciousness plays the *integration role*: it coordinates the functioning of otherwise self-contained cognitive modules, by integrating the information processed by these modules into a central, domain-general stream of information. If some representational state of an agent is not part of such a “global workspace” (to use Baars’ [1997] influential metaphor), then behavior motivated by that state does not satisfy the Control condition, because it is not controlled *by the agent*, as such, but merely by her sub-personal mechanisms. Nor would it satisfy the Identification condition: it is presumably an agent’s person-level beliefs and desires that could count as her established commitments and values. Levy (2014) argues, further, that behavior motivated by a state not part of the global workspace would *eo ipso* fail to satisfy the Reasons-Responsiveness condition as well: if a behavior is not caused by a state that is part of the global workspace, then that behavior reflects only a subset of the agent’s reasons. But for *the agent herself* to exhibit a suitable degree of reasons-responsiveness, her actions need to be responsive to the full range of her person-level mental states.³²

There is, however, room to doubt that consciousness' playing the integration-role entails that intentions governing controlled action need to be (actually or potentially) conscious. First of all, it is not clear whether consciousness *uniquely* plays the integration role. There is certainly conceptual room for a dissociation between integration and phenomenal consciousness; the two concepts are not the same. And there is some evidence of actual dissociation.³³ Second, even if consciousness plays an integration-role, it might not integrate the right kinds of items. For example, King and Carruthers (2012) contend that the contents of the global workspace include mental imagery and inner speech, but never include propositional attitudes; if this is right, then the intentions that govern controlled behavior are never conscious.

Nor is it unanimous that consciousness plays the integration-role; some theories of consciousness fail to attribute such a role to consciousness. For example, according to the "Higher-Order Thought" (or "HOT") theory of consciousness, what it is for a mental state to be conscious is for a subject to have a higher-order mental state directed at the first one. So, even if all person-level states are integrated into the global workspace, HOT-theorists deny that person-level states are therefore *conscious*. Though it is consistent with HOT theory to insist that all person-level representations are potentially conscious (that is, to insist that a subject can have a higher-order thought about all of her person-level representations), the view itself doesn't motivate this requirement. Thus, for a HOT theorist, nothing rules out the possibility that controlled actions can be governed by intentions that the agent is not capable of consciously entertaining.³⁴

There remains the final item on our list of proposed cognitive roles for consciousness, namely, that the cognitive role of consciousness is "a grab-gab." The possibility exists that one or more of the individual functions in the grab-bag is necessarily involved in the formation and/or execution of intentions. For example, David Hodgson (2012) argues that mental states he calls "gestalts" – namely, the binding together of many features in an experiential whole – can only be entertained consciously, and play a crucial role in action-planning.³⁵ Or again, Roy Baumeister (2010) presents evidence that we can only understand certain types of sequentially-structured thoughts consciously, e.g., those that depict logical and causal relations. Paradigms of such sequentially-structured thoughts are *narratives*, and it is only via the mental rehearsal of narratives that we freely choose between alternative action-plans.

There are plenty of other specific functions that consciousness is claimed to uniquely perform: constructing a self-concept; learning new, complex behaviors; building associations between events occurring at a temporal gap from one another; semantically processing sentences; planning sequences of actions, as a means to an end; thinking through a multi-step problem; and so on.³⁶ If any of these functions makes a crucial difference to the formation or execution of the intentions that govern controlled action, then consciousness will turn out to be integral to free action. But it depends on what it is for consciousness to "uniquely" perform these functions – namely, whether consciousness *necessarily* performs them, or whether it is a fluke of our evolutionary history that consciousness always or typically performs them. That human evolution delegated certain functions to consciousness doesn't mean that it had to.³⁷ Granting that humans actually rely on consciousness when they act freely, it does not immediately follow that free action and consciousness are related in any modally strong way.

3 Empirical Considerations

In recent decades, an enormous empirical literature has been growing regarding the relationship between consciousness and action in general, and consciousness and free action in particular. In much of this literature, the empirical findings are understood as threat-posing: they

challenge cherished assumptions about the relationship between our conscious states and our actions. Such challenges can be taken as having either of two implications for a connection between consciousness and free action: (a) a conceptually or metaphysically necessary connection thought to hold between consciousness and free action turns out not to hold after all; or (b) such a connection *does* hold, but fewer of our actions are actually suitably connected to our conscious states than pre-theoretic intuition would suggest (and thus we are less free than we think we are). I will discuss empirical findings from four contexts: pathologies of action; vision-for-action processing; laboratory dissociations; and situational studies.

1. *Pathologies of action.* There are a number of documented pathologies in which a subject's ordinary experience of agency is dramatically inhibited or altered.

- (1) **Anarchic Hand syndrome:** one of an agent's hands engages in apparently goal-directed behavior, sometimes at odds with what she is consciously trying to do, yet the agent is unable to stop or redirect the behavior.³⁸
- (2) **Utilization behavior:** an agent cannot prevent herself from responding habitually to an environmental stimulus.³⁹
- (3) **Delusions of alien control:** an agent experiences that and/or avows that her behaviors are under the control of an external will.⁴⁰
- (4) **Global automatism:** an agent engages in complex behaviors while unconscious (e.g., during sleep, seizure, or coma).⁴¹

In what sense do these cases challenge the idea that free will and consciousness are connected (in concept or in fact)? A natural, first reaction is that they pose no such challenge: if they are cases of action at all, they are cases of frustratingly *unfree* action, precisely because conscious oversight is diminished or absent. But, at the same time, they hint at a challenge regarding the actual extent of the connection. If in pathological cases an agent's hand can engage in complex, goal-directed behavior without or even contrary to the agent's conscious will, then perhaps one's conscious will is causally irrelevant to the behavior of one's hand even in normal cases. Behaviors as simple as picking up a cup and as complex as driving⁴² or composing and sending emails⁴³ can occur in the absence of conscious awareness, driven by non-conscious "action scripts."⁴⁴ Given that such behaviors can occur in the absence of a conscious cause, it is plausible that they occur without being consciously caused even in cases where subjects are consciously aware of performing them.

Now, there is a general consensus that pathological actions are *routinized*, in the sense that they do not express genuine novelty, creativity, or problem-solving on the part of the agent.⁴⁵ They are thus of a piece with habitual actions, as discussed above (such as squeezing the steering wheel or signaling a turn). So, the possibility of pathological action puts pressure on the constitution view and the strong causal-dependence view only to the extent that habitual action does.

2. *Functional dissociation between perceptual and motor processing.* The human visual system serves two primary functions: *acquiring information* about the local environment and *facilitating action* vis-à-vis the local environment. In the last twenty years, neuroscientists have come to believe that the processes that serve these two functions are anatomically and causally isolated from one another to a certain extent.⁴⁶ Evidence has come in the form of pathological dissociations (patients who cannot recognize objects but can act on them normally, and vice versa) as well as dissociations in healthy subjects. For example, Haffenden and Goodale (1998) used plastic discs to create a version of the Ebbinghaus Illusion, where two identical circles appear to differ in size, because one is surrounded by a ring of larger circles and the

other is surrounded by a ring of smaller circles. They found that when subjects use their thumb and index finger to estimate the size of each disk, grip aperture reflects illusory size, whereas when subjects use their thumb and index finger to *reach for* each disk, grip aperture reflects actual size.

Some theorists have interpreted these findings as indicating that a lot of motor action is “zombie action,” i.e. not consciously controlled.⁴⁷ Presumably, when subjects in the Haffenden and Goodale study use their thumb and forefinger to estimate disc size, their estimates correspond to, and are controlled by, their perceptual experience of the discs. But their actual movement vis-à-vis the disks do not correspond to, and so are not controlled by, their perceptual experience of the discs. And there is no reason to think that the illusory nature of perceptual experience in the case in question *creates* the dissociation; rather it *reveals* a dissociation that’s always at play.

But it is at best misleading to characterize subjects’ grip aperture as “zombie action.” While it’s true that *something* is unconsciously controlled when a normal subject picks up an object, that something is not an intentional action but rather a component, aspect, or specification of an intentional action, namely, the precise grip aperture used (and countless other motor details, presumably). But there is a big difference between zombie action in *that* sense and zombie action in the sense of anarchic hand movement. In the former but not the latter case, it makes sense to say that the subject is doing what she consciously intends to do.

What the Haffenden and Goodale study and others like it teach us is that much of the implementation-details of our actions are unconsciously controlled. This point is also familiar from studies of *expert* behavior, which have shown that the better an agent is at some task, the fewer of the details of the means and mechanisms whereby the agent completes the task are consciously experienced.⁴⁸ For example, when a novice pianist plays “Für Elise,” he’ll need to consciously intend to curl his fingers thus and so, to read one measure ahead, etc. An expert pianist may need to consciously intend nothing more than *to play “Für Elise.”* Reaching-behavior is, after all, just another type of (ubiquitous) expert behavior. We expert reachers need consciously intend nothing more than *to pick up the disc*; all of implementation-details are taken care of unconsciously.⁴⁹

Expert action is much like habitual action in that both are typically implemented unconsciously. But whereas habitual action is typically *triggered* unconsciously, expert action is more likely to be triggered consciously – at least under some description (*play the piece, score a point, take a drink*, etc.). That we can intend to perform actions without being consciously aware of the motor-details is not to say that those actions are not governed by conscious intentions, however much those intentions under-specify the motor-details.

3. *Laboratory dissociations.* A lively empirical research program has come together in the last three decades exploring the relationship between physical action and the phenomenology of willing.⁵⁰ Experiments of two sorts have been used to argue that actions are not, or are not always, dependent on agents’ conscious will in the way agents take them to be. First, the work of Benjamin Libet and colleagues suggests that actions can be caused by neural events that temporally precede the agent’s conscious experience of willing them. The experiments that suggest this conclusion involve instructing subjects to spontaneously flex their hand whenever they wish, and then to report the time at which they had the conscious “urge or intention” to do so. (A clock whose “second hand” – actually an arc of illuminated dots – completes a circuit in two and a half seconds is used to mark time.) Meanwhile, measurements of electromagnetic activity of the brain are taken. Libet (1985) found that a particular spike in electrical activity was consistently recorded about 350 microseconds prior to any conscious experience of willing to flex. This pre-motor spike in the electrical output of the

Supplementary Motor Area had been previously discovered and dubbed the “readiness potential.” Libet was able to show that the readiness potential appears not only before motor actions but before any experience of will associated with those actions.⁵¹

What is the significance of this result for the relationship between consciousness and freedom? Many scientists, and a few philosophers, take Libet’s experiments to have shown that we have much less free will than we thought we had. For example, Gregg Caruso has concluded, “Our conscious selves are not in the driver’s seat as we typically believe. This is a devastating blow to our everyday understanding of free will.”⁵² Those like Caruso who draw this skeptical lesson from Libet’s experiments are committed to three claims: first, that the metaphysical connection between consciousness and free action is at least as intimate as that specified by the strong causal-dependence view; second, that Libet’s experiments show that in the cases in question, consciousness plays no causal role in producing action; third, that the cases in question involve the sorts of actions we would pre-theoretically take to be paradigms of free action. But all three of these claims have been questioned. Max Velmans (1991) rejects the first, maintaining instead that from Libet’s experiments we learn that free actions can be unconsciously caused. (Libet has thus provided an empirical corrective to a conceptual error, according to Velmans.) Alfred Mele (2009) rejects the second. All that Libet’s experiments show is that something (the readiness potential) in fact occurs in the brain prior to subjects’ flexing. But this does not by itself establish that readiness potentials are sufficient causes of flexings: (a) consciousness could be contributing an intervening cause between readiness potential and flexing; (b) readiness potentials sometimes occur in the absence of any flexing, as one of Libet’s experiments demonstrates;⁵³ and (c) a conscious, freely formed, *standing* intention to flex-when-one-feels-like-it might be causally necessary.⁵⁴ And the third assumption is rejected by those who note the peculiar *pointlessness* of wrist-flexing: the subject has no reason to flex rather than not to flex at any particular moment.⁵⁵ Such “Buridan’s ass” cases – cases of choosing arbitrarily between two inconsequential options – are not very much like cases for which we take free will to be important: cases of bringing one’s cherished values to bear in deliberating about consequential options, and then directing one’s behavior accordingly. Those who grant the first and second assumptions can still insist that, if Libet has shown that we are less free than we thought we were, he has not shown that we are *very much* less free than we thought we were, or that we are not free when we *care* to be free.

There is a second type of experiment that has been used to argue for a dissociation between consciousness and free action. Daniel Wegner and colleagues have tried to show that subjects can have the experience as of willing an action without performing or even trying to perform that action – that there can be “false positives” among conscious volitional episodes. Two experiments of Wegner’s suggest this possibility. (1) The “I-Spy” experiment:⁵⁶ an experimental subject sits adjacent to a confederate of the experimenter. Both place their hands on a computer mouse that controls a cursor on a visible computer screen. The screen depicts an assortment of objects. The subject is instructed to move the cursor, or to allow the cursor to move, among the depicted objects, finally resting on any one of them after a few seconds. Meanwhile the subject hears music combined with a stream of apparently random words, which in some trials includes the name of an object depicted on the screen, and in some trials does not. For some trials, the confederate is instructed to guide the cursor to the image of the object named in the word-stream. After each trial, the subject is asked to rate the degree to which she or he had intended (vs. merely *allowed*) the cursor to land where it landed. On trials when the confederate guides the cursor to an object named in the word-stream, subjects report a higher degree of intendedness than on similar trials in which the object is not named. (Notably, the naming of a depicted object in the word-stream did not raise the probability that the subject would guide the

cursor to part of the screen where that object is depicted.) (2) The “Helping Hands” experiment:⁵⁷ an experimental subject faces a mirror, while wearing a smock that allows the arms of a confederate (who is standing directly behind the subject) to appear where his or her arms would ordinarily be. The confederate proceeds to perform a series of familiar tasks, such as clapping and making a fist. During some trials, subjects hear instructions that precede and correspond to the actions performed. After these trials, subjects reported a higher degree of intendedness than after those trials during which no instructions were audible.

Wegner (2002) concludes from these experiments, in tandem with others that purport to show the possibility of conscious-volitional “false negatives,”⁵⁸ that the conscious experience of will is always, or usually, an epiphenomenon. This inference is dubious, for a couple of reasons. First, it isn’t clear that in the “false positives” experiments, subjects have had an experience of willing at all. Rather, it seems just as likely that subjects have had no such experience, but are nevertheless inclined to attribute an action to themselves, as we often do in the absence of volitional experience (e.g., a glass is knocked off of a table; before there is any time for me to consciously will to act, I instinctively reach out to catch it – an action I happily attribute to myself), and that such attributions are fallible. Thus, Wegner has not demonstrated that conscious willing is *ever* epiphenomenal. But even if he is right that experimental subjects are having epiphenomenal conscious-volitional episodes in the I-Spy and Helping Hands experiments, his generalization to all or most conscious-volitional episodes is probably not warranted. Wegner seems to be reasoning as follows: in some cases, the experience of conscious willing is non-veridical; we ought to explain action-production in cases of non-veridical conscious willing in the same way as in cases of veridical conscious willing; so conscious willing plays no causal role in either case. But as multiple commentators have pointed out,⁵⁹ just as perceptual hallucinations do not show us that normal perception is illusory, so volitional “hallucinations” do not show us that, in normal cases, our experience of ourselves as conscious agents is illusory.

How do Wegner’s findings bear on the connection between consciousness and freedom? First, if Wegner is right that conscious, volitional episodes of a certain type – specifically, experiences of choosing to execute a bodily movement – *never* generate (let alone constitute) action, then either (a) the constitution view and the strong causal-dependence view are false or (b) one or the other is true and our actions are unfree. But the *weak* causal-dependence view is consistent with epiphenomenalism regarding consciously willing to move one’s body, so long as conscious episodes of a different sort – the conscious cultivation of a habit, the conscious experience of willing an action sometime in the future, etc. – are causally relevant to action-production. Second, if Wegner is right that the experience of willing *sometimes* fails to initiate action in any sense, then this would put pressure on the constitution view. (While advocates of the constitution view do allow that executive consciousness can fail to issue in any bodily movement, Wegner’s claim is stronger: agents can experience themselves as initiating action when they are not even *trying* to initiate action.)

4. *Situational factors.* A challenge of a different kind comes for the voluminous “situationist” literature in psychology. The studies that comprise this literature suggest that human behavior is influenced, in profound, predictable, and irrational ways, by the situations we find ourselves in. Behavior-manipulating factors described in this literature include the following:

- (1) **Bystanders:** The likelihood that a subject offers help to someone in need of medical attention is inversely proportional to the number of other people in the area, and proportional to their level of concern for the emergency.⁶⁰

- (2) **Haste:** The likelihood that a subject offers help to someone in need of medical attention is proportional to how big of a rush he or she is in.⁶¹
- (3) **Luck:** The likelihood that a subject offers help to someone who has dropped his papers is greatly increased if the subject has recently found a coin.⁶²
- (4) **Authority:** Despite expressing strong misgivings, subjects are willing to apply intense electrical shocks to someone if an authority figure is telling them to.⁶³
- (5) **Watching eyes:** Subjects contribute considerably more money to a cause when an image of a pair of watching eyes is nearby.⁶⁴

Despite the differences between the manner and extent of effect that these factors have on human behavior, a few philosophers have taken them to jointly pose a threat to free will, or at least to constrain the extent of our free will. But exactly what this threat comes to is a matter of ongoing debate.⁶⁵ In fact, different situational factors are problematic in different ways, depending on which of the four conditions on freedom is being discussed.⁶⁶

Whether or not the studies mentioned above give us reason to doubt the *extent* of our freedom, do they tell us anything about the relationship between consciousness and freedom? That is, do they demonstrate that consciousness and action are separable in ways that are at odds with the constitution view, the causal-dependence view, or the counterfactual-dependence view? They do not seem to. Consider the constitution and causal-dependence views. The influence of situational factors on human action is consistent with conscious intentions' directly contributing to action. Plausibly, when a subject puts more money in an "honesty box"⁶⁷ than she would have had an image of watching eyes not been posted nearby, or when she more readily helps a stranger than she would have had she not just discovered a dime in a phone booth, these factors influence her precisely *via* consciousness: they influence the degree of her *felt inclination* toward performing some action. So long as it is still in some sense "up to her" whether to follow that inclination, she is still exerting conscious control over her actions.

The counterfactual-dependence view seems to receive some confirmation from the situationist literature. Consider a version of the counterfactual-dependence view according to which an action is free only if (a) the agent can consciously entertain the reasons for which she acts, and (b) can consciously entertain reasons for acting otherwise. When subjects are under the influence of the bystander effect, it's plausible that they act out of a desire to "fit in" with the social group. Yet these same subjects explicitly deny being influenced by the presence and behavior of the other people in the study.⁶⁸ In these circumstances subjects *cannot* consciously entertain (some of) the reasons for which they act, contra clause (a). When subjects are under the influence of the haste effect, it's plausible that they are acting without properly considering the needs of the "medical patient" whose path they cross. In these circumstances subjects cannot consciously entertain (some) reasons for acting otherwise than they do, contra clause (b). The counterfactual-dependence view is thus able to explain in what sense subjects under the influence of bystander and luck effects are less than fully free.

Whether an agent *has the capacity* in a situation to consciously entertain certain reasons is not straightforward; it depends on which aspects of the situation are held fixed. For example, I may be capable of consciously entertaining a reason for acting *if only I would pause for a moment to reflect*, or *if only I were informed about the situationist literature in social psychology*,⁶⁹ or *if only I had received a more rigorous moral education*. Thus the question "Was action A done freely"? will be answerable in many cases only relative to a set of possibilities, and in terms of a notion of freedom that can come in degrees.⁷⁰

I have reviewed four of the most frequently discussed empirical findings related to the connection between consciousness and action, but there are others. Some psychologists have

taken these findings to present a radically revisionary conception of human behavior, one that demotes conscious processing to a mere spectatorial role, and that gives pride of place to unconscious processes.⁷¹ But the growing consensus is that these revisionary conclusions are not warranted by the experimental data.⁷² Moreover, empirical evidence is accumulating that supports the opposite conclusion. Studies on the causal efficacy of consciousness have revealed that consciousness regularly drives action in precisely the ways we intuitively expect, such as via deliberation and planning (Baumeister, Masicampo, and Vohs 2011), and even in moment-to-moment action control, when rival motivations and action-plans present themselves (e.g., when the impulse to drop a hot plate conflicts with the goal of carrying it to the dinner table) (Morsella 2005). Far from undercutting a conception of ourselves as in conscious control of our actions, empirical science seems to be confirming it.

4 Conclusion

Though very few of the considerations we have explored decisively favor one view over its rivals, nevertheless the following two lessons can reasonably be drawn from the foregoing.

First, the possibility of free habitual and expert actions gives us good reason to think that the constitution and strong causal-dependence views are too strong. When philosophers and scientists have taken the empirical discoveries of Libet and Wegner as a threat to free will, they have implicitly assumed that all free actions must be constituted or directly caused by consciousness. Since that assumption is too strong, we can rule out the more revisionary of glosses on Libet and Wegner's work. But this doesn't mean that the constitution and causal-dependence views are without merit. They rest on the intuitive idea that the *direct exercise of control* is a conscious matter. The weak causal dependence view can accommodate this intuition. For example, it could be held that every free action is either a constituted by consciousness (e.g., by being identical to a conscious trying) or is the causal descendent of an action constituted by consciousness. This is a plausible way of understanding what the Control condition requires.

Second, if the Control condition requires that free action is causally dependent on consciousness, the Reasons-Responsiveness condition plausibly requires that free action is counterfactually dependent on consciousness. One strong reason to think so is that it explains how the situationist literature in cognitive psychology exposes limitations to our freedom. Even if our actions trace their causal ancestry to the conscious exercise of agentic control, if we are incapable – at least holding fixed certain aspects of the context – of consciously entertaining the reasons for which we act, then we do not act freely, in the fullest sense. Free will is probably conceptually connected to consciousness in more than one way – causally, in terms of control, and counterfactually, in terms of reasons-responsiveness.⁷³

Notes

- 1 This is a classic, if currently highly contested, view of the epistemic role of consciousness. It goes back at least to Descartes and the British empiricists. See Gertler (2012) for discussion of some recent re-treads.
- 2 John Searle (1991) is responsible for putting this view on the table.
- 3 Bernard Baars' (1997) "Global Workspace Theory" of consciousness is the classic version of this view, but it has been developed in many ways.
- 4 See e.g. Marcel (1988), Hurley (1997), Baumeister, Masicampo and Vohs (2011), and Dehaene (2014).

- 5 Reasons-Responsiveness is sometimes treated as an account of control, and sometimes as a separate condition.
- 6 The debate between compatibilists and incompatibilists turns on whether the Open-Future condition is included in our concept of freedom. I will mention the Open-Future condition only very briefly in what follows; controversies surrounding the other three conditions matter more to whether and how consciousness and freedom are connected. Thus the question of how consciousness and freedom are related is largely independent of the debate between compatibilists and incompatibilists.
- 7 The assertion that phenomenal-consciousness and access-consciousness are close-enough-to-co-extensional for present purposes calls for a more thorough defense than I can give here. As a start: most consciousness-researchers seem to assume that, under most circumstances, mental items are phenomenally conscious if and only if they are available for verbal report. There are two reasons to think so. First, at least until quite recently, researchers in philosophy and psychology overwhelmingly claimed phenomenal consciousness as the target of their inquiry, even as Block (1995) told them they were conflating two notions. Second, excepting the study of unusual psychological effects (blindsight, subthreshold priming, etc.), phenomenal- and access-consciousness are experimentally operationalized in exactly the same way, namely, via verbal report. When a subject is unable to report being aware of some stimulus (or explicitly reports being unaware), this is taken as an indicator that she is *unconscious* of that stimulus, not that it is conscious in some non-phenomenal way.

My own view is that access-consciousness, as Block originally delineated the category, shouldn't be treated as a species of consciousness. Nevertheless, there is a related category of non-phenomenal, occurrent mental states that should be studied alongside phenomenal consciousness. This is the category of the states that are phenomenally conscious or waiting-in-the-wings to become so (e.g., representations of items on a grocery list that one has committed to working memory). I suspect that consciousness-researchers who say they are studying access-consciousness are often studying this phenomenology-*plus* type of consciousness.

- 8 Horgan, Tienson and Graham (2003), Bayne (2008), and Nida-Rümelin (2017).
- 9 That there is irreducibly executive phenomenology is an outlier view in contemporary analytic philosophy and cognitive science, but seems to have been the standard view in nineteenth- and twentieth-century phenomenology. See Kriegel (2015), p. 83. In the eighteenth century Thomas Reid wrote that "We are conscious of making an exertion, sometimes with difficulty, in order to produce certain effects" (1788/1983, p. 332).
- 10 That is, there may be many things that an agent is trying to do by a single action. But only one of these things she is trying to do is identical to her *basic* action, i.e., the action the agent performs without performing any other action as a means to performing it. Danto (1963) first drew attention to the category of basic action.
- 11 For a recent discussion of this "argument from total failure" see Grunbaum (2008). It's noteworthy that David Hume discusses a total-failure case, but reaches a negative conclusion about executive phenomenology: "A man suddenly struck with a palsy in the leg or arm, or who had newly lost these members, frequently endeavours, at first, to move them, and employ them in their usual offices. Here he is as much conscious of power to command such limbs, as a man in perfect health is conscious of power to actuate any member which remains in its natural state and condition. But consciousness never deceives. Consequently, neither in the one case nor in the other, are we ever conscious of any power" (1777/1975, p. 66).
- 12 O'Shaughnessy (1973), p. 370. Hossack (2003) talks of "volitions" rather than "tryings," but seems to have the same phenomenon in mind.
- 13 Compare Schroeter (2004): "I ... suggest that the agent is the conscious self" (p. 642); a conscious self "has the power to send go-signals to initiate the execution of the actions which have been selected" (245). (Note, though, that Schroeter says he is talking about the *access-conscious* self. And whether he endorses the constitution view or merely the causal-dependence view is not clear.)
- 14 O'Shaughnessy (2003), O'Brien (2003), and Hossack (2003) all appeal to versions of the Self-Awareness Argument.

- 15 O'Connor (1995).
- 16 Chalmers and McQueen (Forthcoming) have also developed a wave-collapse theory of mental causation, though not explicitly as an account of libertarian free will.
- 17 Arthur Danto (1985, p. 541) makes much the same point in the context of *projects* rather than *habits*: "A slow-motion film of Matisse shows the artist making countless decisions with his fingers that at normal speed looks like a single confident chalk stroke defining the edge of a leaf. He may or may not have been conscious of each decision ... Consciousness is evolution's gift to us for rather special deliberative employment having to do, as responsibility and free will have to do, with *courses* of action – with *projects*."
- 18 This example comes from Mele (2009).
- 19 From their study of folk conceptions of free action, Stillman, Baumeister, and Mele (2011) conclude that freedom is generally associated with conscious deliberation about, and successful pursuit of, long-term goals. Thus the Indirect Control condition probably better reflects the folk notion of free action than does the Direct Control condition. In keeping with this theme, Nahmias (2002, p. 537) writes, "We all aim to be able, in many types of situations, to act without the feeling of will, but in accord with the plans we have consciously formed. And we feel free and responsible for the actions that accord with these plans, regardless of whether we consciously will each of them."
- 20 Fischer and Ravizza (1998) label (roughly) this type of counterfactualist reasons-responsiveness "guidance control."
- 21 See Searle (1991) p. 56.
- 22 See McGinn (1988), Searle (1991), and Kriegel (2003).
- 23 This problem, sometimes labelled "the disjunction problem," is carefully developed in Fodor (1990). It has close affinities with the "gavagai" problem discussed at length by Quine in his (1960). Fodor thinks the problem can be solved; others are not so optimistic, such as Searle (1991), Adams and Aizawa (1994), and Horgan and Graham (2012).
- 24 Searle (1991), Ludwig (1996) and Strawson (2008) all give versions of this argument.
- 25 As it happens, King and Carruthers (2012) argue that no propositional attitudes are ever conscious, and thus that there probably is no such thing as a "true self" in the sense presupposed by the Identification condition.
- 26 Relatedly, Neil Levy (2014) argues that there is a necessary connection between access-consciousness and moral responsibility: in order to be morally responsible for their actions, agents must be conscious of facts that explain the valence of [their actions'] moral significance" (2014, p. 37). For those who think (as I do, but Levy does not) that an access-conscious state is "conscious" in any meaningful sense only if it is a potentially phenomenally conscious state, Levy's view counts as a version of the counterfactual-dependence view.
- 27 See, e.g. Ruiz-Mirazo and Moreno (2004) and Jones (2015).
- 28 See Jeannerod (1997).
- 29 Schroeter (2004) and Shepherd (2015) both stress this role for consciousness.
- 30 Of course, we hesitate to categorize somnambulant behaviors as free actions. But if the comparator model is right, this can't be because the somnambulist fails to satisfy the Control-condition.
- 31 See Drayson (2014) for a helpful history of the personal/subpersonal distinction in cognitive science.
- 32 It's not clear to me whether Levy is thus committed to the causal-dependence view or the counterfactual-dependence view. That is, it is not clear whether he thinks that behavior that satisfies the Reasons-Responsiveness condition has to issue from *actually* or merely *potentially* globally-available mental states.
- 33 Morsella (2005). Levy (2014) reviews and argues against a number of recent, putative cases of dissociation.
- 34 David Rosenthal, the original and longest-standing advocate of HOT theory, explicitly embraces the epiphenomenality of consciousness when it comes to action-explanation in his (2008). As far as I can tell he is silent on the question of whether controlled action can be guided by intentions the agent is not even capable of consciously entertaining (i.e. whether the counterfactual-dependence view is true).

- 35 Baumeister (2010) seems to concur: “Consciousness is the picture the brain constructs for itself from a (nearly) impossible number of individual pieces of sensory information” (40).
- 36 The literature is voluminous, growing, and rich with controversy. See Marcel (1988), Hurley (1997), Baumeister, Masicampo and Vohs (2011), and Dehaene (2014).
- 37 Polger (2007).
- 38 Marchetti and Della Sala (1998).
- 39 Lhermite (1983).
- 40 Spence (2002).
- 41 Cartwright (2004), Levy and Bayne (2004).
- 42 Penfield (1975).
- 43 Siddiqui, Osuna and Chokroverty (2009).
- 44 See Levy (2014), p. 74ff.
- 45 Levy (2014), Dehaene (2014).
- 46 Milner and Goodale (1995). Milner & Goodale’s original picture was probably too tidy; see Shepherd (2015).
- 47 Cf. Clark (2001), Koch and Crick (2001), Wu (2013).
- 48 Shepherd (2015).
- 49 I am granting, along with the majority of philosophers of cognitive science, that the Haffenden and Goodale study and others like it reveal that there are unconsciously controlled aspects of expert action. See Ian Phillips (2018) for an important dissenting opinion. He argues that the relevant experiments do not warrant even this modest conclusion.
- 50 The essays in Pockett, Banks, and Gallagher (2006) amount to opinionated summaries of much of this research.
- 51 Many Libet-style experiments have been performed in the past three decades, some apparently supporting Libet’s original conclusions, and some calling those conclusions into serious question. Haggard and Eimer (1999) claim to have shown that the experience of willing is regularly preceded not by the readiness potential but by a different event, the “lateralized readiness potential.” Schlegel et al. (2013) partly replicate these findings, but do not think they license Libet’s conclusions: “Neither the RP nor LRP appears to be a correlate of consciousness or conscious volition. The RP may instead be a correlate of general anticipation or preparation or merely of ongoing activity that is neither anticipator nor preparatory” (p. 334) Schurger, Sitt and Dehaene (2012) likewise suggest that the RP be treated as so much neural noise.
- Soon et al.’s (2008, 2013) experiments include a choice between two rival actions – pushing buttons with one or the other index finger (2008); adding vs. subtracting numbers (2013). They find that the readiness potential is not the earliest predictor of which choice subjects will make; as early as 7 seconds prior to motor action (2008) and 4 seconds prior to mental action (2013), neural activity in frontopolar cortex and parietal cortex occurs that can be used to predict which action the subject will perform – but only with about 60% accuracy.
- A number of commentators have from the very beginning raised questions about the design of Libet’s experiments (see peer commentary included in Libet [1985]), but for present purposes I am granting that Libet’s experimental paradigm is basically sound. For much more on Libet-style experiments and their interpretation, see Sinnott-Armstrong and Nadel (2010).
- 52 Cf. Caruso (2012), p. 194. By “understanding of free will” Caruso appears to mean our *beliefs about the extent* of free will, rather than our conception of it.
- 53 In one of the experiments, subjects were instructed to (a) *intend* to flex, but (b) consciously *veto* this intention. Libet reports that “A large RP [readiness potential] preceded the veto, signifying that the subject was indeed *preparing* to act, even though the action was aborted by the subject” (Libet 1999, p. 52). Thus Libet himself rejects an inference from his findings to epiphenomenalism. Now, Mele’s gloss is that Libet’s instructions (a) and (b) cannot be followed simultaneously, and thus that Libet interprets the results incorrectly. Either way, Libet’s data include cases of readiness potentials that are not followed by flexings.

- 54 Eiler and Roessler (2003) make much the same point: “the causally relevant contribution of the subjects’ conscious intentions lies in their prior decision to move their hand “sometime,” delegating the task of deciding exactly when to move to some subpersonal mechanism...Under these circumstances, delegating control to subpersonal mechanisms may be precisely a way of choosing when to act” (42).
- 55 See Breitmeyer (1985), p. 539; Mele (2009), p. 85.
- 56 Wegner and Wheatley (1999).
- 57 Wegner describes the experiment in his (2002), pp. 80–81.
- 58 I.e., cases in which an agent acts but has no conscious experience of doing so. Wegner calls these cases “automatizations.” They include “table-turning,” a practice popular in the nineteenth century in which a group of people lay hands on a table-top which proceeds to rotate, and this motion is attributed to occult forces; and “facilitated communication,” in which a helper “assists” a disabled person to type meaningful prose.
- 59 E.g. Nahmias (2002), Bayne (2006), and Shepherd (2015).
- 60 Latane and Darley (1968).
- 61 Darley and Batson (1973).
- 62 Isen and Levin (1972). This study remains influential despite concerns about replicability. Cf. Blevens and Murphy (1974) and Weyent and Clark (1976).
- 63 Milgram (1974).
- 64 Bateson, Nettle and Roberts (2006).
- 65 See Nelkin (2005), Vargas (2013), and Herdova (2016).
- 66 Nelkin (2005) and Vargas (2013) make a strong case that the central threat is to Reasons-Responsiveness. For example, it is plausible that haste prevents subjects from responding to a need precisely because haste prevents them from being able to engage in the sort of deliberation required (a) to recognize the need as a reason for acting and/or (b) to act for the sake of that reason once recognized. But not all of the factors are troubling for this reason. Luck – e.g., finding a dime in a phone booth – seems actually to boost subjects’ capacity to recognize a need as a reason to act. I think the best way to characterize the perceived threat to freedom from situational luck is that it falsifies the Control-condition: the subject’s decision to help appears to be dictated by her elevated mood, and so no more up to her than other blessings of an elevated mood (increased focus, optimism, etc.). The influence of authority in Stanley Milgram’s experiments is probably best understood as a falsification of the Identification-condition, since subjects appear to act contrary to their best judgment. None of the factors can be seen as a falsification of the Open Future condition, since in the relevant experiments, subjects’ behavior fails to be fully predictable given situational factors.
- 67 That is, the designated cash receptacle for purchased made “on the honor system,” i.e. in the absence of a cashier.
- 68 “Despite the obvious and powerful inhibiting effect of other bystanders, subjects almost invariably claimed that they had paid little or no attention to the reactions of the other people in the room.” (Latané and Darley 1968, p. 220).
- 69 Mele and Shepherd (2013) argue that becoming familiar with situationist literature is a way of increasing one’s degree of free agency.
- 70 Vargas (2013) argues at length for this idea.
- 71 See e.g. Bargh and Ferguson (2000) and Dijksterhuis, Chartrand and Aarts (2007).
- 72 Newell and Shanks (2014), Baumeister, Masicampo and Vohs (2011).
- 73 Thanks to Neil Levy, Sandra Visser, Tim O’Connor, and Lisa Woodward for helpful comments on an earlier draft of this chapter.

Bibliography

- Adams, F. and Aizawa, K. (1994). Fodorian semantics. In: *Mental Representation* (ed. S. Stich and T. Warfield). 223–242, Cambridge, MA: Blackwell.
- Baars, B. (1997). *In the Theater of the Consciousness: The Workspace of the Mind*. Oxford: Oxford University Press.

- Bargh, J. and Ferguson, M. (2000). Beyond behaviorism: on the automaticity of higher mental processes. *Psychological Bulletin* 6: 925–945.
- Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters* 2: 412–414.
- Baumeister, R. (2010). Understanding free will and consciousness on the basis of current research findings in psychology. In: *Free Will and Consciousness: How Might They Work?* (ed. R. Baumeister, A. Mele and K. Vohs), 24–42. Oxford: Oxford University Press.
- Baumeister, R.F., Masicampo, E.J., and Vohs, K.D. (2011). Do conscious thoughts cause behavior? *Annual Review of Psychology* 62: 331–361.
- Bayne, T. (2006). Phenomenology and the feeling of doing: Wegner on the conscious will. *Does consciousness cause behavior*, 1, 169–186.
- Bayne, T. (2008). The phenomenology of agency. *Philosophy Compass* 3: 182–202.
- Blevens, G. and Murphy, T. (1974). Feeling good and helping: further phonebooth findings. *Psychological Reports* 34: 384–388.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18: 227–287.
- Breitmeyer, B. (1985). Problems with the psychophysics of intention. *Behavioral and Brain Sciences* 8 (4): 539–540.
- Cartwright, R. (2004). Sleepwalking violence: a sleep disorder, a legal dilemma, and a psychological challenge. *American Journal of Psychiatry*, 161(7), 1149–1158.
- Caruso, G. (2012). *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*. Lanham, MD: Lexington Books.
- Chalmers, D. and McQueen, K. (Forthcoming). Consciousness and the collapse of the wave function. In: *Consciousness and Quantum Mechanics* (ed. S. Gao). Oxford: Oxford University Press.
- Clark, A. (2001). Visual experience and motor action: are the bonds too tight? *Philosophical Review* 110 (4): 495–519.
- Danto, A. (1963). What we can do. *Journal of Philosophy* 60 (15): 435–445.
- Danto, A. (1985). Consciousness and motor control. *Behavioral and Brain Sciences*, 8(4), 540–541.
- Darley, J. and Batson, C. (1973). From Jerusalem to Jericho: a study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology* 27 (1): 100–108.
- Dehaene, S. (2014). *Consciousness and the Brain*. New York: Penguin Books.
- Dijksterhuis, A., Chartrand, T.L., and Aarts, H. (2007). Effects of priming and perception on social behavior and goal pursuit. In: *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* (ed. J.A. Bargh), 51–132. Psychology Press.
- Drayson, Z. (2014). The personal/subpersonal distinction. *Philosophy Compass* 9 (5): 338–346.
- Eiler, N. and Roessler, J. (2003). Agency and self-awareness: mechanisms and epistemology. In: *Agency and Self-Awareness: Issues in Philosophy and Psychology* (ed. N. Eiler and J. Roessler), 1–47. Oxford: Clarendon Press.
- Fischer, J. and Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Fodor, J. (1990). *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
- Gertler, B. (2012). Renewed acquaintance. In: *Introspection and Consciousness* (ed. D. Smithies and D. Stoljar), 89–123. New York: Oxford University Press.
- Ginet, C. (1990). *On Action*. New York: Cambridge University Press.
- Grunbaum, T. (2008). Trying and the arguments from total failure. *Philosophia* 36: 67–86.
- Haffenden, A. and Goodale, M. (1998). The effect of pictorial illusion on prehension and perception. *Journal of Cognitive Neuroscience* 10: 122–136.
- Haggard, P. and Eimer, M. (1999). On the relation between brain potential and the awareness of voluntary movements. *Experimental Brain Research* 126: 128–133.
- Halvorson, H. (2011). *The Measure of All Things: Quantum Mechanics and the Soul*. In: *The Soul Hypothesis: Investigations into the Existence of the Soul* (ed. M.C. Baker and S. Goetz), 138–167. Continuum Press.
- Herdova, M. (2016). What you don't know can hurt you: situationism, conscious awareness, and control. *Journal of Cognition and Neuroethics* 4: 45–71.

- Hodgson, D. (2012). *Rationality + Consciousness = Free Will*. New York: Oxford University Press.
- Horgan, T. and Graham, G. (2012). Phenomenal intentionality and content determinacy. In: *Prospects for Meaning* (ed. R. Schantz), 321–344. Germany: De Gruyter.
- Horgan, T., Tienson, J., and Graham, G. (2003). The phenomenology of first-person agency. In: *Physicalism and Mental Causation* (ed. S. Walter and H.-D. Heckmann), 323–341. Imprint Academic.
- Hossack, K. (2003). Consciousness in act and action. *Phenomenology and the Cognitive Sciences* 2: 187–203.
- Hume, D. (1777/1975). *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, 3e (ed. L.A. Selby-Bigge). Oxford: Oxford University Press.
- Hurley, S.L. (1997). Nonconceptual self-consciousness and agency: perspective and access. *Communication and Cognition* 30: 207–248.
- Isen, A. and Levin, P. (1972). Effect of feeling good on helping: cookies and kindness. *Journal of Personality and Social Psychology*, 21(3), 384.
- Jeannerod, M. (1997). *The Cognitive Neuroscience of Action*. Oxford: Blackwell.
- Jones, D. (2015). Agency through autonomy: self-producing systems and the prospect of biocompatibilism. *Journal of Cognition and Neuroethics* 3 (1): 217–228.
- King, M. and Carruthers, P. (2012). Moral responsibility and consciousness. *Journal of Moral Philosophy* 9: 200–228.
- Koch, C. and Crick, F. (2001). On the Zombie within. *Nature* 411 (6840): 893–893.
- Kriegel, U. (2003). Is intentionality dependent upon consciousness? *Philosophical Studies* 116 (3): 271–307.
- Kriegel, U. (2015). *The Varieties of Consciousness*. New York: Oxford University Press.
- Latane, B. and Darley, J. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology* 10 (3): 215–221.
- Levy, N. (2014). *Consciousness and Moral Responsibility*. New York: Oxford University Press.
- Levy, N. and Bayne, T. (2004). Doing without deliberation: automatism, automaticity, and moral accountability. *International Review of Psychiatry* 16: 209–215.
- Lhermitte, F. (1983). 'Utilization behavior' and its relation to lesions of the frontal lobes. *Brain* 106 (2): 237–255.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* 8 (4): 529–566.
- Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies* 6 (8–9): 47–57.
- Ludwig, K. (1996). Explaining why things look the way they do. In: *Perception* (ed. K. Akins), 18–60. New York: Oxford University Press.
- Marcel, A. (1988). Phenomenal experience and functionalism. In: *Consciousness in Contemporary Science* (ed. A. Marcel and E. Bisiach), 121–158. New York: Oxford University Press.
- Marchetti, C. and Della Sala, S. (1998). Disentangling the alien and anarchic hand. *Cognitive Neuropsychiatry* 3: 191–207.
- McGinn, C. (1988). Consciousness and content. *Proceedings of the British Academy* 74: 225–245.
- Mele, A. (2009). *Effective Intentions*. New York: Oxford University Press.
- Mele, A. and Shepherd, J. (2013). Situationism and agency. *Journal of Practical Ethics* 1 (1): 62–83.
- Milgram, S. (1974). *Obedience to Authority: An Experimental View*. New York: Harper & Row.
- Milner, D. and Goodale, M. (1995). *The Visual Brain in Action* Oxford: Oxford University Press.
- Morsella, E. (2005). The function of phenomenal states: supramodular interaction theory. *Psychological Review* 4: 1000–1021.
- Nahmias, E. (2002). When consciousness matters: a critical review of Daniel Wegner's *The Illusion of Conscious Will*. *Philosophical Psychology* 15: 527–541.
- Nelkin, D.K. (2005). Freedom, responsibility and the challenge of situationism. *Midwest Studies in Philosophy* 29 (1): 181–206.
- Newell, B., and Shanks, D. (2014). Unconscious influences on decision making: A critical review. *Behavioral and brain sciences*, 37(1), 1–19.
- Nida-Rümelin, M. (2017). Freedom and the phenomenology of agency. *Erkenntnis* (83): 1–27.

- O'Brien, L. (2003). On knowing one's own actions. In: *Agency and Self-Awareness: Issues in Philosophy and Psychology* (ed. J. Roessler and N. Eilan), 358–382. New York: Clarendon Press.
- O'Connor, T. (1995). *Agents, Causes, and Events: Essays on Indeterminism and Free Will*. New York: Oxford University Press.
- O'Shaugnessy, B. (1973). Trying (as the mental 'pineal gland'). *Journal of Philosophy* 70 (13): 365–386.
- O'Shaugnessy, B. (2003). The epistemology of physical action. In: *Agency and Self-Awareness: Issues in Philosophy and Psychology* (ed. J. Roessler and N. Eilan). New York: Clarendon Press.
- Penfield, W. (1975). *The Mystery of the Mind*. Princeton, NJ: Princeton University Press.
- Phillips, I. (2018). Unconscious perception reconsidered. *Analytic Philosophy* 59 (4): 471–514.
- Pockett, S., Banks, W.P., and Gallagher, S. (2006). *Does Consciousness Cause Behavior?* MIT.
- Polger, T. (2007). Rethinking the evolution of consciousness. In: *Companion to Consciousness* (ed. S. Schneider and M. Velmans), 77–92. Oxford: Blackwell.
- Quine, W. (1960). *Word & Object*. Cambridge, MA: MIT.
- Reid, T. (1788/1983). *Thomas Reid's Inquiries and Essays* (ed. R.E. Beanblossom and K. Lehrer). Indianapolis, IN: Hackett.
- Rosenthal, D. (2008). Consciousness and its function. *Neuropsychologia* 46: 829–840.
- Ruiz-Mirazo, K. and Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life* 10 (3): 235–259.
- Schlegel, A. et al. (2013). Barking up the wrong tree: readiness potential reflect processes independent of conscious will. *Experimental Brain Research* 229: 329–335.
- Schroeter, F. (2004). Endorsement and autonomous agency. *Philosophy and Phenomenological Research* 69: 633–659.
- Schurger, A., Sitt, J.D., and Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences* 109: E2904–2913.
- Searle, J. (1991). Consciousness, unconsciousness and intentionality. *Philosophical Issues* 1 (1): 45–66.
- Shepherd, J. (2015). Conscious control over action. *Mind and Language* 30 (3): 320–344.
- Siddiqui, F., Osuna, E., and Chokroverty, S. (2009). Writing emails as part of sleepwalking after increase in zolpidem. *Sleep Medicine* 10: 262–264.
- Sinnott-Armstrong, W. and Nadel, L. (2010). *Conscious Will and Responsibility: A Tribute to Benjamin Libet*. New York: Oxford University Press.
- Smithies, D. (2012). The mental lives of zombies. *Philosophical Perspectives* 26 (1): 343–372.
- Soon, X. et al. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience* 11 (5): 543–545.
- Soon, X. et al. (2013). Predicting free choices for abstract intentions. *Psychological and Cognitive Sciences* 110 (15): 6217–6222.
- Spence, S. (2002). Alien motor phenomena: A window on to agency. *Cognitive Neuropsychiatry* 7 (3): 211–220.
- Stapp, H. (2005). Quantum interactive dualism: an alternative to materialism. *Journal of Consciousness Studies* 12: 43–58.
- Stillman, T.F., Baumeister, R.F., and Mele, A.R. (2011). Free will in everyday life: autobiographical accounts of free and unfree actions. *Philosophical Psychology* 24: 381–394.
- Strawson, G. (2008). *Real Materialism and Other Essays*. New York: Oxford University Press.
- Vargas, M. (2013). Situationism and moral responsibility: free will in fragments. In: *Decomposing the Will* (eds. T. Vierkant, J. Kiverstein and A. Clark), 325–350. New York: Oxford University Press.
- Velmans, M. (1991). Is human information processing conscious? *Behavioral and Brain Sciences* 14 (4): 651–669.
- Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT.
- Wegner, D. and Wheatley, T. (1999). Apparent mental causation: sources of the experience of will. *American Psychologist* 54: 480–492.
- Weyant, J. and Clark, R. (1976). Dimes and helping: the other side of the coin. *Personality and Social Psychology Journal* 3: 107–110.
- Wu, W. (2013). The case for zombie agency. *Mind* 122 (485): 217–230.