

CONSCIOUSNESS AND MORAL RESPONSIBILITY: SKEPTICAL CHALLENGES AND THEOLOGICAL REFLECTIONS

by *Aku Olavi Visala* 

Abstract. Some philosophers and scientists have argued that we humans cannot be held morally responsible for anything. Invoking results of the neurosciences and the cognitive sciences, they argue that humans lack the kind of conscious control and awareness required for moral responsibility. For theological ethics and Christian theology as a whole, moral responsibility is indispensable. I will begin by outlining some empirical results that are invoked in support of moral responsibility skepticism. I will, then, examine the subsequent discussion and the question why conscious awareness is central to moral responsibility. Consciousness contributes to morally relevant control over action in multiple ways. I will briefly examine some accounts of conscious control that are resistant to the skeptical challenge. Although the empirical results might lead us to revise the degree and range of conscious control, there seems to be enough of it to ground many everyday practices of responsibility. I will conclude the article with some theological reflections.

Keywords: cognitive science; consciousness; free will; moral responsibility; neuroscience; psychology; self

Recently, some philosophers and scientists have argued that we humans cannot be held morally responsible for anything. Invoking recent results of the neurosciences and the cognitive sciences, they argue that (most) humans lack the kind of conscious control over action (“free will”) that moral responsibility requires. It seems to us that our actions are, most of the time, brought about by our conscious intentions and reasons that have led to the formation of those intentions. The scientific free will skeptics argue that this “seeming” is, in fact, false—a kind of illusion. In reality, our sense of conscious control does not track the actual causes of our actions. It follows from this, the argument continues, that humans seldom—or never—fulfill the conditions of moral responsibility.¹

Aku Olavi Visala is currently a Docent in Philosophy of Religion and Research Fellow of the Finnish Academy, Faculty of Theology, University of Helsinki, Helsinki, Finland; e-mail: aku.visala@helsinki.fi.

[*Zygon*, vol. 56, no. 3 (September 2021)]

www.wileyonlinelibrary.com/journal/zygon

© 2021 The Authors. *Zygon*® published by Wiley Periodicals LLC on behalf of Joint Publication Board of *Zygon* ISSN 0591-2385
This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Skepticism toward conscious control over action can be motivated by both scientific and more philosophical considerations. In this article, I will focus on scientifically motivated skepticism (although the empirical and the philosophical case for skepticism are not completely distinct).² Skeptical worries have generated an enormous amount of philosophical and scientific literature in the last two decades.³ My aim is to introduce the core issues to a theological readership in order to invite theologians to examine the psychological assumptions behind their accounts of moral responsibility more carefully. Although contemporary philosophical theologians have been eager to talk about various metaphysical and doctrinal issues surrounding free will (grace, sin, providence, divine determinism), they have said very little about concrete psychological issues revolving around consciousness and control.

I will begin by outlining some empirical results that are invoked in support of free will skepticism. I will then examine some responses to these challenges and look at the role of consciousness in action control in detail. One possible response is to adopt an account of moral responsibility that foregoes the need for conscious control of action altogether. Although this is an interesting proposal, I will present some arguments against it and go on to discuss the other option, that is, the possibility of conscious control even while acknowledging the thoroughgoing influence of nonconscious mental mechanisms in action generation. To this I will dedicate the bulk of this article. Finally, I will conclude with some Augustinian theological reflections.

EMPIRICAL EVIDENCE FOR EPIPHENOMENALISM

I take it to be uncontroversial that most accounts of Christian religious and moral life entail that individuals have significant direct and indirect control over their moral and spiritual lives. This control is a presupposition for much that constitutes such a life. Christians blame and praise one another; they assess each other's moral characters, and invite each other to account for their actions and omissions. For religious life especially, both intersubjective and intrasubjective moral appraisals have crucial importance. A person's relationship to herself often consists of various self-appraisals that guide her moral life. Emotions and attitudes, such as regret and guilt as well as self-gratitude and praise, are significant for a person's self-image and for the regulation of her moral life. Moral appraisals are also at home in religious social lives. Intersubjective attitudes of love and forgiveness are held in high esteem by Christian ethics. In addition, moral and spiritual life involves an aspect of development and change. Sanctification is a process of purification and moral development. A person is supposed to become more aware of her sins, regret and repent them, and develop virtues instead of vices. Here, the Christian is invited to develop her moral awareness of

the significance of her actions as well as her capacities for self-control and self-reflection (see, e.g., Zahl 2020).

One central function of a theological account of moral responsibility is to make sense of the practices and attitudes that constitute moral and spiritual life, as well as to show that they are properly justified or grounded. It is obvious that in order to account for these practices and attitudes, we must assume that a person has some measure of conscious control over her actions and internal mental life. If it turned out that humans overall were incapable of controlling their actions in a morally significant way, it would be difficult to see how, for instance, making moral appraisals, guilt, and blame could be justified at all. Moreover, if it turned out that humans lacked the necessary mental resources to shape their internal mental life or moral character, it would render at least some accounts of virtue-based moral and spiritual life untenable. This would also apply to any account of sanctification that requires some measure of cooperation between the individual and God. This is why theologians must engage with the nitty-gritty details of human moral psychology, instead of just developing accounts of free will and moral responsibility in “the metaphysical sphere” and discussing issues like divine determinism and God’s omniscience.

A number of philosophers and scientists have recently argued for free will skepticism on the grounds that conscious mental processes are epiphenomenal with respect to actions (Caruso 2012). We humans seem to have very little conscious control over our actions, because we lack significant awareness of those factors that, in fact, cause our behavior. We have no control over most of our action-generating mental mechanisms, because they operate outside conscious awareness. For early defenders of these arguments, like psychologist Daniel Wegner (2002), the results imply that our conscious will is an illusion. If we further assume that a person cannot be held morally responsible for an action that she did not control, it seems that moral responsibility is threatened.

The target of the skeptic is basic desert moral responsibility. The core idea of just deserts is that a person deserves to be blamed or praised simply on the basis of the moral worth of her actions. Philosopher Derk Pereboom expounds the idea:

For an agent to be morally responsible for an action in this sense is for it to be hers in such a way that she would deserve to be blamed if she understood that it was morally wrong, and she would deserve to be praised if she understood that it was morally exemplary. The desert at issue here is basic in the sense that the agent would deserve to be blamed or praised just because she has performed the action, given an understanding of its moral status, and not, for example, merely by virtue of consequentialist or contractualist considerations. (Pereboom 2014, 2)

The reason why some empirical results threaten moral responsibility is that they seem to suggest that when a person forms an intention to act,

the person is unaware of the causes that lead to the formation of that intention. The subsequent action might be caused by the intention, but the intention is not caused by the person's conscious reasons, but by something else entirely. Her consciously accessible reasons, desires, and values are ultimately epiphenomenal with respect to the action. If this is the case, those values that the person is aware of and which she endorses, might be mostly irrelevant with respect to her actions. As a consequence, they cannot function in the role of explaining or otherwise making sense of her actions. Epiphenomenalism threatens moral responsibility, because it suggests that the actual causal chain that leads to an action bypasses the person's consciously endorsed reasons to act. But it is the person's values, reasons and intentions that are supposed to ground our moral appraisals of her. If her actions do not "flow" from her reasons, intentions, and values, it does not make sense to blame the person for her actions.

Let us say that an agent commits a morally wrong act while sleepwalking and would not act in such a way, if awake. Suppose also that the state of sleepwalking is such that the person is not aware of what she is doing. Now, it would seem unjustified to blame the agent for that action, because the agent lacks control over her actions. While sleepwalking, the agent is neither aware of the moral significance of her actions nor are her values and reasons connected to the morally wrong action. If the agent is blamed for the action, it is open to her to offer various excuses: she can complain that the blame is unfair, since she did not intend to commit the action; she did not know what she was doing. Had she had conscious control, she would have acted otherwise. The skeptic insists that most people most of the time are in an analogous situation: they lack conscious control over their actions, so any attribution of basic desert moral responsibility is unjustified.

A central assumption of the debate is that basic desert responsibility requires the fulfillment of two conditions: the "free will condition" or control condition and the epistemic condition.⁴ The skeptical argument attacks both simultaneously: people do not know what they are doing and they lack conscious control over it. This reveals the crucial role that consciousness plays in both of these conditions. If a person has very little or no conscious awareness of her mental life that leads to an action (reasons, intentions), she cannot be said to control the action. Moreover, if a person lacks conscious awareness of the morally significant features of an action, the person does not know what she is doing.

One significant source of evidence for the skeptical claim is the neuroscience of conscious volition and decision-making. Most famously, neuroscientist Benjamin Libet conducted a series of experiments about how conscious decisions to act are temporally related to measurable brain events. Without going into the details, Libet-style experiments involve the measuring of brain activity in some way (originally by electroencephalography, EEG, subsequently by functional magnetic resonance imaging,

fMRI) and its timing in relation to the conscious decision to act. The act measured in the experiments is usually a basic motor movement, like flexing of one's hand, such as in the original Libet (1985) experiment. The surprising finding in the original study was that there is an activation of the motor cortex 400 milliseconds before the subject became aware of her conscious decision to act. Later, many others, like neuroscientist John-Dylan Haynes (2013) and his colleagues, have advanced the basic paradigm and presented further evidence as to how simple decisions of test subjects can be predicted from fMRI data seconds before the subjects become conscious of them (Soon et al. 2008). One skeptical interpretation of these results is that the conscious decision to act is not the cause of the action, but rather an effect of the triggering of the motor neurons, which takes place prior to any conscious awareness. Before any conscious forming of an intention, the brain is already preparing to perform the action. So, the conscious intention to act cannot be the cause of the action, but rather a side effect of brain mechanisms working outside conscious awareness.

For critics of free will, the Libet paradigm demonstrates how our conscious awareness of forming an intention to act can come apart from the actual causes of actions. Wegner (2002), in particular, has argued that our sense of acting intentionally is not the cause of our actions, but an illusion created by nonconscious mental processes. Wegner and his research group devised many experiments that were able to induce a sense of ownership over actions in subjects, while the subjects actually had no such ownership over their actions. Wegner also describes experiments that purport to show that subjects can be put in conditions where they in fact initiate actions but lack the conscious awareness of bringing that action about. Similarly, a sense of intentional action can be induced in a test subject, when the subject is not really acting at all. These experiments raise multiple questions about the phenomenology of free action and the nature of conscious will, which have been extensively discussed in the literature (Nahmias 2002). Wegner himself took the evidence to show that we humans are never in a position, where we consciously control our actions. Instead, the sources of our actions are independent of our conscious selves.

Many philosophers take the cognitive sciences as a whole to provide evidence against conscious control over actions. If we have learned anything from cognitive psychology, it is that most mental processes work outside our conscious awareness. There is a significant amount of evidence for pervasive automatism in the human cognition, or so the skeptical argument goes (Kihlstrom 2008). Particularly well-known are the various hypotheses about the modular structure of human cognition. According to original modularity hypothesis of Jerry Fodor (1983), many cognitive systems, especially those that categorize and process sensory information,

are modular in nature, that is, their processes are fixed, isolated from one another, and their inputs and outputs are not affected by the central cognition. Their workings are what Fodor calls doxastically impenetrable, namely, permanently outside conscious access.

Later, many cognitive scientists have defended the massive modularity thesis, according to which all mental processes are more or less modular in this sense (Barrett & Kurzban 2006). An ardent defender of massive modularity, Peter Carruthers (2006), argues that conscious will is an illusion, because our awareness of our mental states floats free of the actual causes of our actions. Conscious volitions and intentions are not the causes of our actions, but rather *post hoc* products of interpretive mechanisms, stories that integrate our actions and self-image, as it were. According to Carruthers (2006, 408), "... if it can be shown that the only form of access that we have to our own intentions and decisions is interpretive—in this respect like the access that we have to the intentions and decisions of other people—then there is no such thing as conscious willing or conscious deciding."

Carruthers (2006, 406–13) argues that this is in fact the case, because of the way in which our knowledge of our mental states is generated. According to him, our own understanding of the contents of our minds is based on our cognitive capacity to understand other people. We humans automatically and habitually interpret and explain the actions of other people by attributing them various mental states, like beliefs, intentions, and desires. This same exact process is how we attribute beliefs, desires, and intentions to ourselves. However, these mind-reading modules have no access to those modules that generate our actions. They are interpretations that do not track the true causes of our actions (Carruthers 2009; Gazzaniga 2011).

In addition, automatism is also supported by social psychology, where it has long been acknowledged that the individual's behavior often depends on nonconscious social cues. This is especially pertinent in situationist views of moral behavior. According to situationism, the bulk of our moral behavior is to be explained by immediate environmental causes rather than factors involving our internal features, like personality, intentions, and moral character. Social psychologists have created a number of ingenious experimental settings, like Stanley Milgram's (1974) obedience to authority experiments and the Stanford Prison experiment (Haney, Banks, Zimbardo 1973) in order to show the extensive, nonconscious influence of the social context, especially in the case of moral action. These studies highlight the role of implicit social roles and cues in the regulation of moral behavior. Given the truth of situationism, it would seem that our behavior in many morally significant situations would be guided by nonconscious cues and factors of the situation rather than our own consciously endorsed commitments and attitudes (Nelkin 2005). On the situationist

reading of the results, these implicit cues are the main causes of moral behavior in specific circumstances, not the character, values, and beliefs of the individual. The link between our commitments, values and intentions and our actions is crucial for moral appraisal. If they are mostly irrelevant in explaining an individual's moral behavior in a given situation, it seems that moral responsibility is severely undermined.

NO CONSCIOUSNESS NEEDED FOR RESPONSIBILITY?

One way of avoiding the skeptical challenge would be to insist that conscious control of action is not required for moral responsibility after all. There are two ways doing this. First, one could argue that while control is required for moral responsibility, one does not need that control to be conscious. In other words, an account of control can be provided that does not include conscious access to relevant mental states. This would neatly bypass the skeptical challenge. Second, one could forego the whole notion of control altogether and ground moral responsibility on something else. Next, I will briefly consider these possibilities. I will begin from the second option.

In contemporary literature, control-based accounts of moral responsibility have competitors that look impervious to the skeptical challenge. In what follows, I will stick to Neil Levy's (2014a, 11–12) practice and call these expressivist accounts of moral responsibility. The core idea is that moral responsibility is grounded in the relationship between the agent's attitudes, judgments and evaluations and her actions (Wolf 1990; Arpaly 2002; Smith 2005). An agent can be held morally responsible for an action if that action expresses or represents the evaluations, attitudes, and judgments of the agent. Notice, that such features of the agent can be involuntary and outside the control of the agent.

In what follows, I will examine the expressivist account of George Sher (2009), who has directly discussed the issue of consciousness and responsibility. Sher criticizes what he calls the searchlight view of moral responsibility. According to this view,

an agent's responsibility extends on as far as his awareness of what he is doing. He is responsible only for those acts he consciously chooses to perform, only for those omissions he consciously chooses to allow, and only for those outcomes he consciously chooses to bring about. (Sher 2009, 4)

Sher argues that the searchlight view is false and puts forward a number of everyday cases of moral appraisal that seem to speak against it. If the searchlight view were true, it would rule out our responsibility over actions that were caused by ignorance, lack of imagination or poor judgment. However, we habitually hold people responsible precisely for such lapses of attention and judgment.

Sher mentions a famous situation in Quentin Tarantino's *Pulp Fiction* (1994), where Vincent (played by John Travolta) accidentally shoots another person, who was sitting at the backseat of his car. Vincent is brandishing a loaded weapon around and, in his agitation, accidentally fires it with head-splattering result. This lands Vincent and his hitman colleague Jules in deep trouble. It is obvious Vincent did not intend to fire the weapon and perhaps was also ignorant that it was loaded. Given the searchlight view, Vincent would not be responsible for shooting of his passenger. If moral responsibility requires conscious control, Vincent does not seem to be responsible. However, it seems rather obvious that while Vincent's action was not consciously intended or controlled, it was a result of his poor judgment and lapse of awareness. He should have been more careful—should have known better. As a consequence, it seems obvious that Vincent cannot completely escape blame in this situation.

Sher goes on to develop an alternative to the searchlight view, along the lines of expressivist theory of moral responsibility. Sher argues that an agent can be held responsible for those actions that flow out of the agent's deeply held judgments and beliefs. This is the reason why such accounts are sometimes called deep self or real self—accounts in the literature. The deep self of the agent is constituted by a nexus of mental features (intentions, evaluations, and tendencies) and their causal histories. Most of these features are involuntary and outside the agent's control. Moreover, in many cases, the agent herself is not aware of these features, but her actions are nevertheless brought about by such features. The person can justifiably be held responsible for actions flowing out of these features even if the person is ignorant of them, because those actions have their roots in the person's core psychology.⁵

Sher's account would explain neatly why Vincent is blameworthy for his misfiring. Although Vincent did not intend to fire his gun, there is a deep-seated failure in Vincent's psychology that renders him responsible for the shooting. Vincent should have checked whether his gun was loaded and he should have been more careful. He failed to do so, because of his easily aggravated and careless tendencies that are very much a central part of who he is. Given that Vincent is a hitman and gangster by profession, Vincent's ignorance expresses his judgments about the value of other people's lives (especially those of idiots). Vincent is clearly ignorant of these features of himself, but that ignorance does not excuse him, because that ignorance can be traced back to his deep-seated judgments that constitute the kind of person he is.

An expressivist account of moral responsibility seems resistant to skeptical challenges, at least initially. Sher would not be bothered by Libet-style experiments, since these experiments only measure the subject's awareness of his conscious initiation of action. Furthermore, an expressivist account

would not be challenged if it turned out that humans lacked conscious awareness of many of their reasons, attitudes and even intentions driving their actions. Indeed, some defend expressivist accounts of responsibility precisely because they consider them less susceptible to skeptical challenges, unlike control-based accounts (Arpaly 2002).

In the context of this article, I cannot go into the debate about whether the expressivist account is a plausible theory of moral responsibility. Nevertheless, I will present some reasons why I find it as an unsatisfactory response to the skeptical challenge. First, there is experimental evidence on folk intuitions about moral responsibility that speak against the expressivist view. In a set of experiments, Joshua Shepherd (2012) asked subjects to react to vignettes that described nonconscious behavior and those describing conscious behavior. One of the key results was that the subjects associated free will and moral responsibility with behavior that was accompanied by consciousness rather than nonconscious mechanisms. Shepherd also conducted experiments to find out whether the folk view corresponds to the expressivist theory. The results were surprising. Shepherd (2015a, 938) concludes, “the right interpretation, in my view, is that consciousness is central to folk views of free and responsible action, and that the way in which it is central is not captured by extant Deep Self Views.”

Another defect of the expressivist account is that, as Gregg Caruso and others have argued (Caruso and Morris 2017), it is not deep enough to ground the kind of moral responsibility that is at stake in the skeptical argument. Recently, philosophers have distinguished different aspects of moral responsibility (Shoemaker 2011). *Attributability* is at the core of the expressivist argument. An action can be attributed to an agent, when the action can be taken as an expression of the agent’s deeply held judgments and evaluations. Such an action is something that the agent owns. For the expressivist, attributions of this kind are enough to ground a wide range of responsibility attitudes, like blame and praise. However, Caruso (2018) has suggested that while the expressivist view might be able to ground some measure of moral responsibility in the sense of *attributability*, this might not be enough for the kind of *basic desert* moral responsibility described above. There are powerful arguments in favor of the view that *basic desert* responsibility requires not just *attributability* but some measure of *accountability*. In addition to attribution, the agent is also required to be accountable for that action in the sense that she can be expected to provide justification for them (McKenna 2012). Providing justifications clearly requires conscious access to one’s intentions and reasons for acting. If consciousness-involving accountability of this kind is necessary for *basic desert* responsibility, then the expressivist account cannot ground *basic desert*.

Finally, there is some doubt whether the expressivist view can manage without consciousness even in its own terms. Contrary to Sher, Levy argues (2014a, 88–90) that in order for the agent's actions to express the agent's judgments and evaluations, conscious mental processes are needed to mediate between the agent's actions and judgments and values. Without conscious access, the Deep Self of the person is not connected to the person's actions in such a way as to allow for moral evaluation.

Imagine now that our sleepwalker stabs someone to death. To some extent, the sleepwalker can be responsive to the environment: he grasps the right tool and chooses a target. However, the person's Deep Self does not drive the stabbing. Rather, there are basic motor processes that are responsive to the physical features of the environment, but the responsiveness to the person's values is missing, because conscious awareness of the situation is missing. What consciousness does is that it makes morally relevant factors of the situation (the victim is innocent, for instance) such that they can interact with the person's Deep Self. Without conscious awareness, moral features of the situation "do not interact with the broad range of attitudes constitutive of his evaluative agency. They do not interact with his personal-level concerns, beliefs, commitments, or goals" (Levy 2014a, 89). So, in order for the actions of an agent to express her deeply held judgments, goals, and beliefs, the agent must be, at least at some point, consciously aware of having such attitudes.

The discussion on expressivist views teaches us an important lesson. Sher and others are clearly right about something: "if agents must be conscious of all the mental states that shape their behavior, no one would ever be responsible for anything" (Levy 2014a, 36). The skeptic is surely correct in claiming that our conscious access to our mental life and the moral features of various situations is limited and spotty at best. Those who seek to respond to the skeptical challenge must, therefore, show that there can be moral responsibility grounding control over action despite the evidence of spotty conscious access.

CONTROL AFTER THE SKEPTICAL CHALLENGE

The discussion around the skeptical challenge is vast and I cannot even begin to cover it all here. However, I will highlight some features of the debate that open the door for the possibility of conscious control over action.

First, one rather direct way to respond to the skeptical challenge is to point out that there is plenty of evidence suggesting that conscious cognitive processes feature prominently in explanations of many kinds of behaviors. Social psychologist Roy Baumeister and colleagues (2011, 351) conclude, "the evidence for conscious causation of behavior is profound, extensive, adaptive, multifaceted, and empirically strong."

Although Baumeister and colleagues admit that we often perform actions that are not consciously initiated, this says nothing about the significant contributions that consciousness makes with respect to behavior in other ways.

There is research suggesting that conscious processes provide the human cognition the ability to simulate future states of affairs, plan for various contingencies, control some intuitive responses (especially emotional responses) and plan for the future. Particularly relevant for our case are results about long-term implementation intentions, which “translate general, abstract intentions into specific behavioral plans, of the form “If X happens, then I will do Y.” Dozens of careful studies have confirmed that these cause changes in behavior over and above merely intending, desiring, goal setting, and valuing” (Baumeister, Masicampo and Vohs 2011, 336; Gollwitzer and Sheeran 2006).

There is also evidence for conscious control in the sense people can modify and override automatic responses. Oftentimes, emotions directly trigger actions. However, when a person consciously reflects upon the emotion and its possible consequences, the person can thereby prevent himself from acting upon the emotion. Similarly, there is evidence that acting on negative emotions can be prevented by prompts from other people. Moreover, consciousness contributes to self-control in many ways (Mele 2020). This can be shown by experimentally impairing conscious reflection, which usually results in a situation, where the person’s desires (that have not been endorsed) take over more easily. It seems that at least some automatic processes can be modified, if their workings are brought to conscious awareness. Baumeister has developed a model of “willpower” or self-control that is based on such results. His “energy model of self-control” has been the subject of much debate (Baumeister, Tice, Vohs 2018). Regardless of the fate of the energy model, research on self-control is progressing and there is significant evidence that persons can consciously, albeit indirectly, control their internal mental life (like emotions) as well as control their actions via intentionally shaping their environment.

This last point can be utilized as a response to the problem raised by situationism. Although there are many implicit social cues that drive behavior, “consciousness seems to reduce the power and influence of many of these situational influences” (Baumeister, Masicampo and Vohs 2011, 350). We could follow Manuel Vargas (2013) and turn the whole situation upside down. Rather than take social psychology experiments to undermine conscious control over action, we might argue that such experiments show the opposite. When people learn about such experiments and become aware of their tendencies to submit to authority or blindly follow socially prescribed roles, the effect of these context-dependent tendencies can be significantly mitigated.

One critical response to this suggestion is that the situational influences on behavior are so numerous that no one is able to be aware of such influences, when entering into a morally relevant situation. An average person cannot be expected to know all the relevant psychological studies. Not to mention to be able to take all their results into account when acting.⁶ These considerations show that direct control over situational influences might be, in many cases, very difficult. They do not show, however, the impossibility of indirect control. When a person becomes aware of, say, a tendency to submit to authority or to avoid helping people for morally irrelevant reasons, the person can develop various skills and tendencies to resist such influences. So, we could say that increasing knowledge about implicit social cues and other previously nonconscious factors influencing our behavior can actually enhance the degree of control we exercise over our actions, not simply reduce it.

Let us now move on to the neuroscientific experiments emerging from the Libet paradigm. Enormous amounts of ink have been spilled over this issue (Brass, Furstenberg, Mele 2019). Many philosophers have expressed their doubts about whether such experiments can show the overall lack of conscious control over behavior.

Levy (2014a, 20–24) argues that the results of Libet-style studies and Wegner's studies are—despite initial appearances—irrelevant for assessing whether people are morally responsible. The problem with studies in the Libet paradigm is that they measure the timing of the subject's becoming aware of her intention to act. But why would we think that immediate awareness of one's proximate intention to act would be necessary for control over action? In many cases of stereotypically free and responsible action, direct proximate intentions are missing. Consider, for instance, driving a car. A competent driver acts in multiple ways at once: watching traffic, coordinating bodily movements of hands, legs and responding to various rules and cues. None of these require conscious initiation or even conscious awareness when they are performed. Nevertheless, there is indirect conscious control: the driver can become aware of the various components of driving and modify them, if needed. Furthermore, driving a car is also a stereotypical example of intentional action: there is an overall intention that the agent is consciously aware of having, but the details of the implementation (motor coordination, perception, feedback mechanisms) are performed automatically.

Alfred Mele (2009) makes a distinction between distal and proximal intentions: proximal intentions are decisions to act immediately and distal intentions are longer term commitments to achieving some goal. He argues that in Libet experiments, the subjects have formed a distal intention to participate in the experiment and flex their hands at some point. The concrete implementation (proximal intention) of that flexing is automatized. So, even if we admitted that no brain correlate of a conscious

intention was found in Libet-style studies, it would not rule out there being an overall intention to act (distal intention).

Many critics have suggested that Libet experiments deal only with morally irrelevant actions, like flexing one's hand (Levy 2005). These are a far cry from morally relevant decisions that require deliberation. One cannot generalize the results of such studies over all actions, including complex moral actions. Moreover, there is some evidence to suggest that the readiness potential measured in Libet experiments does not emerge in cases of complex moral decisions (Maoz et al. 2019). Levy also points out that there is no consensus as to what the readiness potential is and what it represents. Mele (2009), for instance, suggests that it could simply be something that prepares the brain to form intentions. Some studies suggest that the readiness potential does not perform the function that Libet attributes to it (Alexander et al. 2015). These points undermine the skeptical argument based on Libet studies.

Moving onto research on automatism, Levy argues that Wegner has not been able to demonstrate that conscious awareness of acting and the actual act are disconnected. Although various illusions of conscious acting can be induced in specific experimental setting, this does not show that conscious awareness is always dissociated from acting. Levy takes visual perception as an example. There are cases of disorder (like blindsight) where the awareness of perceiving something comes apart from actually perceiving something. Clearly,

it would be a mistake to think that actual perception and taking oneself to perceive are not very reliably, and indeed causally, linked. A double dissociation is good evidence that there is not an exceptionless link between the items dissociated, in either direction, but it is not good evidence that there is no direct causal link between them. (Levy 2014a, 19)

So, it seems that while Wegner is able to demonstrate that our awareness of acting and actually acting might come apart, he has not shown that there is a nonconscious mechanism that explains both in all cases (see, also Nahmias 2002). Levy (2014a, 24) concludes, "It is very likely to be false that we are always, or even often, conscious of the precise moment at which a decision is made. But there is little reason to think that this fact threatens moral responsibility."

At this point, we should make some important distinctions about the scope of the skeptical argument. One question is the range of the skeptical challenge: is it supposed to apply to all human actions? Some interpretations of the Libet paradigm and some of Wegner's claims imply a universal range. In other words, all actions are such that conscious mental states contribute nothing to them causally. Many have been skeptical about this strong, universal claim. It became apparent above that there might be many morally relevant actions that fall outside the range of the

skeptical attack in this sense. As a universal thesis based on experimental results, the skeptical claim is not very plausible. Another question is the degree to which conscious control plays a part in explaining actions. Again, the skeptical thesis can be interpreted categorically: there is no conscious control over action whatsoever. As such, the above results and arguments suggest that this categorical interpretation is likely false.

However, even if the universal range is denied and there is evidence that in some cases the degree of conscious control is higher than zero, the skeptical argument retains some of its power. Shepard (2015b, 330–31) is spot on when he writes, “the general claim is that the causal impact is surprisingly insignificant. This is a claim about the degree of consciousness’s involvement, and given the current state of knowledge, replies will be about the degree of consciousness’s involvement.” So, the more plausible (and still moral responsibility undermining) skeptical challenge is not that conscious mental states are epiphenomenal with respect to all actions, but rather that the range of consciously controlled action and the degree of control are both surprisingly small—or at least smaller than we tend to think (Caruso 2015). The conclusion of the skeptical argument is not that humans are incapable of moral responsibility *per se*, but rather that the argument forces us to introduce significant revisions to our moral responsibility practices. Surprisingly, many actions fall outside the range of responsible action completely and the degree of conscious control we have over them is smaller than we think.

CONSCIOUSNESS AND THE EPISTEMIC CONDITION

When the conclusion of the skeptical argument is formulated in the more plausible way, as indicated above, the debate becomes more difficult. One should classify the range of actions to which the skeptical conclusion is supposed to apply and the degree to which conscious control is lacking. This task cannot be taken up in one article. However, let me briefly examine some accounts of conscious control that are compatible with empirical results and retain a significant degree of control over the range of many kind of actions, especially over many everyday moral actions.⁷ According to these views, the fact that our conscious access to our action-generating factors is spotty at best is still compatible with morally significant degree of control.

We must consider the notion of consciousness itself. The skeptical argument derives its power from findings that suggest that conscious mental states have very little to do with action. But what is the kind of consciousness that is in doubt here? The concept of consciousness can be understood in many ways that are not necessarily relevant for moral responsibility. The sleepwalker example above suggests that the sleepwalker lacks creature consciousness, namely, that the person is not

awake at that time. Another notion is phenomenal consciousness, which refers to one possible feature of some mental states, namely, that they involve a specific “feel” or experienced quality as to what it is like to be in that state (Chalmers 1995). Given the controversial nature of phenomenal consciousness, I will remain agnostic as to whether a person is required to be capable of phenomenal consciousness in order to be morally responsible.

However, it seems that the skeptical argument is not after creature consciousness or phenomenal consciousness. The sleepwalker lacks another type of consciousness: she seems to lack access to her mental states. On the crucial question of how this access should be understood, there is disagreement in the literature. A prominent defender of the skeptical argument, Gregg Caruso (2012) adopts a form of the higher order thought (HOT) theory of consciousness, while one critic Neil Levy (2014a) defends a global workspace theory of consciousness. Fortunately, the dissimilarities between these two accounts of consciousness need not bother us too much here. As far as I see it, they both converge on the central claim that the cognitive function of consciousness is that of integration of information from many modular subsystems and broadcasting integrated representations so that modular subsystems can have access to them. It is this function that makes consciousness central for both the control and epistemic side of responsibility.

On the HOT, an agent is access conscious (actual or dispositional) of some mental state when she can represent herself as being in that state. The subsequent representations are called higher order representations (Caruso 2012, chapter 5.3). Crucially for our purposes, these higher order representations are such that they are available for HOT, like the process of decision-making, weighting of options and forming of intentions and plans. The benefit that access consciousness provides to a modular mind is that it allows for the flexible integration of outputs from multiple systems and the broadcasting such representations across systems that would otherwise be distinct. Formed in these terms, the skeptical challenge is this: it seems that most causes of human behavior, both mental and external, are such that they are not accessible by those systems that form higher order representations. Similarly, the content that plays a role in higher order representations plays only a small or insignificant role in the causation of behavior.

Levy thinks that this limited access is not an ultimate threat to moral responsibility.⁸ There are plausible accounts of control over action that are compatible with limited conscious awareness of mental contents and that there is, in most cases of moral action, the capacity to gain access to relevant knowledge in order to ground moral appraisals. With respect to the latter—epistemic criterion—Levy argues that moral appraisals of the person can be grounded if the person has conscious access to those features

of the situation that explains “its moral valence” (Levy 2014a, 37). To be responsible for an action, the person must have access to those features that make the action morally good or bad. What the agent needs to be aware of are facts that explain the moral significance of a particular action, like the badness of theft or the goodness of giving to charity. Such features need to be personally available (accessible) to the agent effortlessly and in such a way as to be used in online reasoning (Levy 2014a, 33). This awareness is indicated by the ability to report the content, namely, to explain it, when asked. Availability of this kind also makes it possible for the content to be widely broadcasted across multiple cognitive systems.

Levy argues that this kind of conscious access is central in explaining many cases of moral behavior, which involves moral reflection, decision-making and responsiveness to shared norms. The skeptic overplays her hand, when she claims that many or most moral actions are accounted for by nonconscious mechanisms. Behavior driven by nonconscious mechanisms can be complicated and responsive to triggering features of the environment, but it is nevertheless too inflexible and insensitive to explain the width of flexible, culturally sensitive agency that most instances of human moral behavior exhibit. To explain this, we must—according to Levy—postulate a wide range of behaviors that have consciousness as a significant component. As I already mentioned above, Levy adopts the global workspace theory, which was originally formulated by Bernard Baars (2002) but has been later developed by Stanislas Dehaene and others (2011). Dehaene and others identify consciousness as network of a distributed set of neurons that connects, collects, integrates, and broadcasts information across various independent systems. For Levy, the most crucial function is that of global broadcasting. Without consciousness, the outputs of individual mind/brain mechanisms would not come together and be available to other mechanisms to use as input.

Levy’s response to the skeptic is, then, that while the skeptic might be able to explain some actions, like flexing of hands or various illusions of conscious acting, by invoking nonconscious mechanisms, it does not follow that complex moral behaviors can be explained such. Furthermore, such results do not show that people overall lack conscious access to relevant features of most moral situations. The situationist challenge, of course, reveals how pervasive implicit cues are and how much they can shape moral behavior. However, it does not show that people overall lack access to morally relevant features of the situation or that the degree of control over actions in such situations is so small, that people should be excused.⁹

Levy (2014b) also discusses complex cases, where the agent is generally conscious but lacks awareness of some specific aspect of her actions. Many skeptics of moral responsibility have referred to the extensive experimental literature on implicit bias, like sexism and racism in hiring practices,

for instance, Uhlmann and Cohen (2005). Consider a person who hires someone less qualified for a job, because of an implicit bias to view male candidates more qualified than female candidates. In these cases, it is typical that the male candidate appears more competent than the female one despite objective criteria.

It would be tempting to say that such a person is liable to blame, given the expressivist account of responsibility: indeed, it seems that such an action would express the person's deeply held attitudes and judgments about males and females and thus qualify as a stereotypically responsible action. However, Levy (2014a, 92–97) argues that this is not the case. Although the hiring decision of the person is clearly sexist, the cognitive bias at work in this case is outside the person's awareness. Let us suppose that the person does not consciously endorse sexism and would reject it, if given the opportunity. Indeed, this is often the case with implicit biases. The reason why the person would not be morally responsible in the case of unjust hiring is that he was unaware of his actual selection criterion (the bias) and instead operated on the basis that the male was indeed more competent than the female. It was this (false) information that was globally broadcast. Because of this, the person's deeply held attitudes and evaluations were interacting with false information and, therefore, not expressed in the person's behavior. Indeed, Levy (2014b) suggests that implicit attitudes in general are not robust enough to ground moral attributions.

CONSCIOUSNESS AND CONTROL

We can now return to the link between consciousness and control. According to a very basic notion of control, by controlling an action “agents deploy behavior in the service of motivational states” (Shepard 2014, 396), like intentions. So, a person controls an action, when that action falls under the content of the intention and that intention plays a causal role in producing the action. Following Alfred Mele, Joshua Shepard suggests that the content of an intention is a plan: an intention includes a representation of what the end state will look like. Control is the agent's ability to succeed in implementing the intention via action. We should also note that, since success comes in degrees, control also comes in degrees. An agent can succeed to implement the intention to a high degree or a low degree. Like Levy, Shepard (2015b, 328–29) also argues that many low-level features of actions are explained by automatic motor systems that control the implementation of intentions, which are higher up in the hierarchy of control. Conscious control does not extend to such functions. However, conscious intentions can make a critical causal contribution to a wide range of actions. The control is, for the most part, executive rather than implementational, that is, conscious intentions provide the overall,

coarse-grained goal of the action, whereas initiation and implementation are, for the most part, automatic.

Let us return to the case of a person driving a car. Levy (2014a, 74–75) explains the automaticity involved by postulating the existence of action-scripts. Action-scripts are sets of motor representations that direct individual motor actions and are responsive to various environmental cues. In most cases, action-scripts are learned through conscious attention and training, but having learned them, the agent need not consciously attend to them. Driving involves a set of such action-scripts, like changing gears and the complex process of steering a car. Normally, when we drive, we trust the action-scripts to do their work without conscious awareness. However, if something unusual and unexpected happens, conscious awareness steps in and begins to deliberate between various action-scripts. So, “conscious processing is needed for the flexible modulation and the interruption, or concatenation, of action scripts. Only when the agent is conscious do a very broad range of internally and externally generated cues modulate behavior” (Levy 2014a, 79).

This kind of conscious control is not threatened by Libet-paradigm studies or arguments about extensive automatism in human cognition. It is also compatible with the fact that the person has spotty access to the factors that cause her action in a given moment. Nevertheless, it leaves intact a degree of control that allows for moral responsibility.

Let us consider one specific account of control-based moral responsibility.¹⁰ One of the most discussed accounts in the literature is that of John Martin Fischer and Mark Ravizza (1998). For Fischer and Ravizza, what matters for moral responsibility is that the action has its roots in the person’s cognition. This happens when the person’s action producing mechanisms are sensitive to a variety of reasons (reasons-responsiveness) and those mechanisms properly belonging to the person (mechanism ownership). When the agent’s actions are products of reasons-responsive systems that belong to the agent, the agent can be said to exercise guidance control over that action. Guidance control of this kind is distinct from regulative control, which refers to the agent’s ability to act otherwise. It is also distinct from notions ultimate control (Kane 1996) prevalent in the literature. The exercise of guidance control, for Fischer, is the minimum threshold for responsibility.

One might initially think that guidance control requires no conscious awareness at all. It seems possible that an agent might act in a reasons-responsive way without being consciously aware of doing so. In his discussion on the issue, Levy argues that despite the appearances guidance control indeed requires significant input from consciousness. “We exercise guidance control over our actions if we would recognize reasons, including moral reasons, as reasons to do otherwise, and we would actually do otherwise in response to some such reason in a counterfactual

scenario” (Levy 2014a, 109). According to Fischer, it is not enough to be minimally reasons-responsive, like the sleepwalker mentioned above, who is clearly responsive to some features of the environment. The sleepwalker is in control of her actions to some degree, but this degree is so low as to allow for moral responsibility. The sleepwalker’s action scripts are not regularly responsive to moral reasons; instead, they are responsive to only a very narrow set of external cues. Responsibility-level sensitivity requires responsiveness to a larger set of factors, especially moral reasons. Without conscious access to morally relevant features of situations, moral reasons are not integrated and broadcasted across cognitive systems.

Another feature of guidance control that requires conscious awareness is that of mechanism ownership. Crucial for the reasons-responsive account is the notion of taking responsibility, or mechanism ownership. As Pereboom already noted above, morally responsible actions must belong to the person performing them. The person must have ownership over them. For Fischer and Ravizza, most psychological mechanisms and tendencies are not the person’s own doing. However, the person can take responsibility over them and their effects by forming true beliefs about their functions and integrating them into her view of herself as an actor.

As far as I can see, the reasons-responsive account of control is quite resistant to the skeptical challenges presented above. First, the reasons-responsive account acknowledges that control comes in degrees, since reasons-responsiveness also comes in degrees. The empirical results, especially those of Libet and Wegner, do not undermine the possibility of moderate reasons-responsive action in most everyday moral situations. Second, the skeptical results do not undermine the possibility of mechanism ownership. As we already noted, situationism clearly shows that taking responsibility for some moral tendencies in our moral behavior might turn out to be rather difficult. In this sense, the empirical results should make us more careful in making moral judgments of people’s actions in complex social situation. However, this difficulty does not rule out the possibility of at least moderate-degree of control via taking responsibility and developing practices and skills to overcome biases and tendencies.

SOME AUGUSTINIAN REFLECTIONS

Let us now summarize the discussion above. First, the results of psychology and cognitive science point toward the fact that most human mental processes have significant nonconscious components. However, despite the skeptical worries, there are reasons to think that a significant degree of conscious control is still possible over a wide range of human moral actions. Notice, that my defense includes the compatibilist theory of control of Fischer and Ravizza. Most skeptics will criticize this account of control, but the arguments would not be—for the most part—scientific or empirical.

I would also like to emphasize that in order to respond to the empirical skeptical challenge thoroughly, one should examine various types of moral actions in multiple contexts in order to determine the range and degree to which nonconscious factors play a part in generating actions. In this sense, the skeptical challenge has not been completely dispelled here. It might turn out that the range of nonconscious influence is so broad and the degree in which nonconscious systems feature in some actions is so high that we must revise some of our responsibility practices (Caruso 2015, 2018) to an uncomfortable degree. However, I do maintain that the thesis of universal epiphenomenalism is hardly supported by the empirical evidence.

Let me conclude with some theological reflections on these results. One interesting feature of both Sher's and Levy's views is what they imply about the acting self. They imply that the acting self cannot be identified with those features that are consciously accessible at a given time. The responsible agent is constituted by a much larger set of facts that can be accessed from the first-person perspective. For Levy (2014a, ix), the conscious self is just a tiny sliver of the totality of mental life. For Sher (2009, 121–22), the self includes the causal features of the body, environment and history that uphold the agent's judgments, beliefs, and other mental features.

Such an account of the nature of the responsible agent fits naturally into an Augustinian framework of the self and the will. In this framework, the agent is always dependent on history, context, and other people in ways, which go beyond the agent's control. We are not self-determining or self-defining creatures, but creatures that mostly find ourselves having certain goals, desires, and characters. Our wills are not our own making. According to Jesse Couenhoven (2017, 38), Augustine does not see the will as a distinct faculty of the person's soul, separate from reason. Instead,

One's *uoluntas* seems to reflect on the whole person in a way that a merely passing whim or fancy—in spite of the loves, desires, or wishes that those inevitably involve—does not. In a deep sense, your *uoluntas* is who you are as a person.

One does not have to be aware of one's ultimate desires, loves and deeply held judgments in order for those to be constitutive of the one's core self. Not only do these features often escape conscious awareness, they also escape conscious control. When Augustine claims that the will of the person is not under her control, he means that the overall outputs of the person's reason, moral judgment and appraisal come outside conscious awareness and control. Even if one acquires conscious access to some of these outputs of one's mental apparatus, that still does not entail control over them.

Another theologically interesting suggestion can be distilled from the discussions above. It seems that the agent's experiences associated with being the source of her actions (the sense of initiating an action, the sense

of ownership of that action) are not necessarily veridical. This claim could have a number of potential theological applications. It could be invoked to explain situations where divine agency and the experience of human agency overlap. Our commonsense assumption is that if another agent in fact causes my action, I lack the appropriate experience of being the source of my action. When this intuition is applied to God, it seems that God's actions and my agency are at odds.

Consider a practical case as an illustration. In 428, Augustine wrote a pastoral letter of advice to a man named Firmus, who was struggling with whether he should become a Christian or not.¹¹ Firmus has apparently read a significant amount of Christian literature and is inclined to take baptism. However, he feels he is not as convinced about the Christian truth as he should be in order to sign up. Inspired by Augustine's earlier texts, Firmus thinks that only God can change a person's will from unbelief to faith. As Augustine himself had many times argued, coming to faith is ultimately something that God does, not the human being. Therefore, it is up to God to bring it about that the person consciously wills to do good and accept baptism. Since Firmus does not experience such an act of God, he decides that it is better to wait for it than to take matters at his own hands, as it were.

Augustine's response to this worry is interesting for our topic. He says that a person's conscious decision to exercise her will in uncertain conditions is compatible with the fact that God ultimately causes her to act. Augustine explains:

Do not wait until he wills it (*nec expectes quando uelit*), as if you were going to offend him if you willed it first. For, whenever you have willed it, you will be willing it with his help and by his working. His mercy, of course, anticipates you so that you may will it, but when you will it, you yourself certainly will it. For, if we do not will when we will, then he does not give us anything when he makes us will.¹²

Given the above discussion on the experience of conscious willing, we might put Augustine's point in the following way: there is no incompatibility between the agent consciously settling a difficult process of decision-making by "just deciding" and the fact that such a decision is also caused by God. In other words, when God brings about an intention in an agent, that agent does not necessarily experience God's actions as extrinsic or herself as simply a passive object of God's actions. Instead, the agent's own "settling of her will" whereby she forms an intention to act can also be brought about by God. Even if one does not first experience the "settledness" or "firmness" of an intention (i.e., one does not experience one's will as settled with the option of accepting baptism) and settles it, this process can still be a result of God's will. Such options should be interesting to explore further.

Finally, a modest account of control, such as the reasons-responsive account of Fischer and Ravizza, is a good fit with Augustinian theology, precisely because of its inherent compatibilism. Augustinian theologians are often forced to make sense of human moral responsibility in the light of divine determinism. Although Couenhoven (2013) has argued that Augustine himself comes close to contemporary expressivist accounts, Michael Preciado (2019) has recently argued that Reformed views of moral responsibility, especially that of Jonathan Edwards', resemble more closely the control-based account of Fischer and Ravizza than anything else. Adopting something like the reason-responsive account would provide the divine determinist an explanation of how humans can be responsible under divine determinism as well as a way to respond to the skeptical challenges arising from the sciences.¹³

NOTES

1. Free will skeptics often disagree as to how this conclusion affects our moral responsibility practices. Some suggest that because we lack conscious control over our actions, we should radically reject or revise our practices, like blaming, punishing, and praising (e.g., Waller 2011). Others hold that some moral responsibility practices can be salvaged: perhaps holding one another responsible has some significant benefits for the human community as a whole that justify such practices even without free will (Shaw, Caruso, and Pereboom 2019).

2. For philosophical cases for free will skepticism, see Pereboom (2014), Waller (2011), and Levy (2012). See also Pereboom and Caruso (2018).

3. See, for example, Baumeister, Mele, and Vohs (2010); Pockett, Bank, and Callagher (2006); Clark, Kiverstein, Vierkant and (2013); and Sinnott-Armstrong (2014). For introduction, see, for example, Caruso (2018) and Nahmias (2010).

4. A number of textbooks confirm this. See, for example, Timpe (2008) and McKenna and Pereboom (2016).

5. Here, Sher is following the lead of other philosophers, who accept the expressivist view of moral responsibility, especially Smith (2005). However, Sher differs from Smith and others by adding something to the picture: not only is the responsible self-constituted by practical judgments, but also by the causal processes that are necessary for the existence for the deliberative process of practical judgments. See Sher (2009, 128–33).

6. I want to thank the anonymous reviewer for this point.

7. I want to point out that I am bypassing many of the philosophical debates about free will skepticism here. In what follows, I will invoke accounts of moral responsibility that are compatibilist in nature. Many skeptics like Caruso (2012), Pereboom (2014), Levy (2012), and Bruce Waller (2011) have multiple philosophical arguments against such accounts, which I do not assess here. They maintain that a correct account of moral responsibility must be an incompatibilist one. A complete response to the skeptical challenge would have to respond to various arguments from manipulation, luck, and determinism as well. This article restricts itself to the empirical skeptical challenge only.

8. Notice, that Levy is a free will and moral responsibility skeptic himself. However, he does not think that skepticism is supported by empirical results about the lack of conscious access.

9. For Levy's response to Carruthers and others invoking the fallibility of our self-knowledge, see Levy (2014b).

10. There are also accounts of nonconscious control in the literature (e.g., Suhler & Churchland 2009).

11. My discussion of this topic is indebted to Timo Nisula's unpublished paper "Nec expectes quando uelit (ep. 2*, 7)—Augustine's persuasions on *voluntas*." The English quotations of Augustine are Nisula's translations (2017).

12. ep. 2*, 7: “nec expectes quando uelit, quasi offensurus eum si ante tu uelis, cum ipso adiuuante atque operante uelis, quandocumque uolueris. praeuenit quidem te misericordia eius, ut uelis, sed cum uoles, tuitique uoles. nam si nos non uolumus, quando uolumus, non ergo nobis aliquid ille confert, cum efficit ut uelimus.”

13. The Academy of Finland graciously funded the research for this article. I would also like to thank the two anonymous referees for invaluable comments.

REFERENCES

- Alexander, Prescott, Alexander Schlegel, Walter Sinnott-Armstrong, Adina Roskies, Peter Ulric Tse, and Talia Wheatley. 2015. “Dissecting the Readiness Potential: An Investigation of the Relationship between Readiness Potentials, Conscious Willing, and Action.” In *Surrounding Free Will: Philosophy, Psychology, Neuroscience*, edited by Alfred Mele, 203–30. Oxford: Oxford University Press.
- Arpaly, Nomy. 2002. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford: Oxford University Press.
- Baars, Bernard. 2002. “The Conscious Access Hypothesis: Origins and Recent Evidence.” *Trends in Cognitive Sciences* 6:47–52.
- Baumeister, Roy, Alfred Mele, and Kathleen Vohs, eds. 2010. *Free Will and Consciousness: How Might They Work?*. Oxford: Oxford University Press.
- Baumeister, Roy, E. J. Masicampo, and Kathleen D. Vohs. 2011. “Do Conscious Thoughts Cause Behavior?” *Annual Review of Psychology* 62 (1): 331–61.
- Baumeister, Roy, Dianne Tice, and Kathleen Vohs. 2018. “The Strength Model of Self-Regulation: Conclusions from the Second Decade of Willpower Research.” *Perspectives on Psychological Science* 13 (2): 141–15.
- Barrett, Clark, and Robert Kurzban. 2006. “Modularity in Cognition: Framing the Debate.” *Psychological Review* 113:628–47.
- Brass, Marcel, Ariel Furstenberg, and Alfred Mele. 2019. “Why Neuroscience Does Not Disprove Free Will.” *Neuroscience and Biobehavioral Reviews* 102:251–63.
- Carruthers, Peter. 2006. *The Architecture of the Mind*. Oxford: Oxford University Press.
- . 2009. “How We Know Our Own Minds: The Relationship Between Mindreading and Metacognition.” *Behavioral and Brain Sciences* 32:121–38.
- Caruso, Gregg. 2012. *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*. Lanham, MD: Lexington Books.
- . 2015. “If Consciousness is Necessary for Moral Responsibility, then People are Less Responsible than We Think.” *Journal of Consciousness Studies* 22 (7–8): 49–60.
- . 2018. Consciousness, Free Will, and Moral Responsibility. *The Routledge Handbook of Consciousness*, edited by Rocco Gennero, 78–91. London: Routledge.
- Caruso, Gregg, and Steven Morris. 2017. “Compatibilism and Retributive Desert Moral Responsibility: On What is of Central Philosophical and Practical Importance.” *Erkenntnis* 82: 837–55.
- Clark, Andy, Julian Kiverstein, and Tillman Vierkant, eds. 2013. *Decomposing the Will*. Oxford: Oxford University Press.
- Chalmers, David. 1995. “Facing up to the Hard Problem of Consciousness.” *Journal of Consciousness Studies* 2 (3): 200–19.
- Couenhoven, Jesse. 2013. *Stricken by Sin, Cured by Christ: Agency, Necessity, and Culpability in Augustinian Theology*. Oxford: Oxford University Press.
- . 2017. “Augustine’s Moral Psychology.” *Augustinian Studies* 48 (1–2): 23–44
- Dehaene, Stanislas, Jean-Pierre Changeux, and Lionel Naccache. 2011. “The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architecture to Clinical Applications.” In *Characterizing Consciousness: From Cognition to the Clinic?* edited by Stanislas Dehaene, and Y. Christen, 55–84. Berlin: Springer.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Fodor, Jerry. 1983. *The Modularity of the Mind*. Cambridge, MA: MIT Press.
- Gazzaniga, Michael. 2011. *Who’s in Charge? Free Will and the Science of the Brain*. New York: Harper Collins.

- Gollwitzer, Philip, and Paschal Sheeran. 2006. "Implementation Intentions and Goal Achievement: A Meta-analysis of Effects and Processes." *Advances in Experimental Social Psychology* 38:69–119.
- Haney, Craig, Curtis Banks, and Philip Zimbardo. 1973. "Interpersonal dynamics of a simulated prison." *International Journal of Criminology and Penology* 1:69–97.
- Haynes, John-Dylan. 2013. "Beyond Libet: Long-Term Prediction of Free Choices from Neuroimaging Signals." In *Decomposing the Will*, edited by Andy Clark, Julian Kiverstein, and Tillman Vierkant, 60–72. Oxford: Oxford University Press.
- Kane, Robert. 1996. *The Significance of Free Will*. Oxford: Oxford University Press.
- Kihlstrom, Joshua. 2008. "The Automaticity Juggernaut—or Are We Automaton after All?" In *Are We Free? Psychology and Free Will*, edited by John Baer, James Kaufman, and Roy Baumeister, 155–80. Oxford: Oxford University Press.
- Levy, Neil. 2005. "Libet's Impossible Demand." *Journal of Consciousness Studies* 12 (12): 67–76.
- . 2014a. *Consciousness and Moral Responsibility*. Oxford: Oxford University Press.
- . 2014b. "Consciousness, Implicit Attitudes, and Moral Responsibility." *Nous* 48 (1): 21–40.
- . 2012. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford: Oxford University Press.
- Libet, Benjamin. 1985. "Unconscious Cerebral Initiatives and the Role of Conscious Will in Voluntary Action." *Behavioral and Brain Sciences* 8:529–66.
- Maoz, Uri, and Gideon Yaffe, and Christof Koch, and Liad Mudrik. 2019. "Neural Precursors of Decisions that Matter: An ERP Study of Deliberate and Arbitrary Choice." *eLife* 8:e39787.
- McKenna, Michael. 2012. *Conversation and Responsibility*. Oxford: Oxford University Press.
- McKenna, Michael, and Derk Pereboom. 2016. *Free Will: A Contemporary Introduction*. London: Routledge.
- Mele, Alfred. 2009. *Effective Intentions: The Power of Conscious Will*. Oxford: Oxford University Press.
- Mele, Alfred, ed. 2020. *Surrounding Self-Control*. Oxford: Oxford University Press.
- Milgram, Stanley. 1974. *Obedience to Authority*. New York: Harper and Row.
- Nahmias, Eddy. 2002. "When Consciousness Matters: A Critical Review of Daniel Wegner's The Illusion of Conscious Will." *Philosophical Psychology* 15 (4): 527–41.
- . 2010. "Scientific Challenges to Free Will." In *A Companion to the Philosophy of Action*, edited by Timothy O'Connor, and Constantine Sandis, 345–56. London: Blackwell.
- Nelkin, Dana. 2005. "Freedom, Responsibility, and the Challenge of Situationism." *Midwest Studies in Philosophy* 29 (1): 181–206.
- Nisula, Timo. 2017. "Nec expectes quando uelit (ep. 2*, 7): Augustine's Persuasions on Voluntarism." Unpublished Presentation at the Helsinki Analytic Theology Workshop.
- Pereboom, Derk. 2014. *Free Will, Agency, and the Meaning of Life*. Oxford: Oxford University Press.
- Pereboom, Derk, and Gregg D. Caruso. 2018. "Hard-Incompatibilist Existentialism: Neuroscience, Punishment, and Meaning in Life." In *Neuroexistentialism: Meaning, Morals, and Purpose in the Age of Neuroscience*, edited by Gregg D. Caruso, and Owen Flanagan, 193–222. Oxford: Oxford University Press.
- Pockett, Susan, William Banks, and Shaun Callagher, eds. 2006. *Does Consciousness Cause Behavior?* Cambridge, MA: MIT Press.
- Preciado, Michael. 2019. *A Reformed View of Freedom*. Eugene, OR: Pickwick.
- Shaw, Elizabeth, David Caruso, and Derk Pereboom, eds. 2019. *Free Will Skepticism in Law and Society: Challenging Retributive Justice*. Oxford: Oxford University Press.
- Shepherd, Joshua. 2012. "Free Will and Consciousness: Experimental Studies." *Consciousness and Cognition* 21:915–27.
- . 2014. "The Contours of Control." *Philosophical Studies* 170:395–411.
- . 2015a. "Consciousness, Free Will, and Moral Responsibility: Taking the Folk Seriously." *Philosophical Psychology* 28 (7): 929–46.
- . 2015b. "Conscious Control over Action." *Mind & Language* 30:320–44.
- Sher, George. 2009. *Who Knew? Responsibility Without Awareness*. Oxford: Oxford University Press.

- Shoemaker, David. 2011. "Accountability, Attributability, and Accountability: Toward a Wider Theory of Moral Responsibility." *Ethics* 121:602–32.
- Sinnott-Armstrong, Walter, ed. 2014. *Moral Psychology: Free Will and Moral Responsibility*, vol. 4. Cambridge, MA: MIT Press.
- Smith, Angela. 2005. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115:236–71.
- Soon, Chun Siang, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes. 2008. "Unconscious Determinants of Free Decisions in the Human Brain." *Nature Neuroscience* 11:543–45.
- Suhler, Christopher, and Patricia Churchland. 2009. "Control: Conscious and Otherwise." *Trends in Cognitive Sciences* 13 (8): 341–47.
- Timpe, Kevin. 2008. *Free Will: Sourcehood and Its Alternatives*. London: Bloomsbury.
- Uhlmann, E., and Cohen, G. 2005. "Constructed Criteria: Redefining Merit to Justify Discrimination." *Psychological Science* 16:474–80.
- Vargas, Manuel. 2013. "Situationism and Moral Responsibility: Free Will in Fragments." In *Decomposing the Will*, edited by Andy Clark, Julian Kiverstein, and Tillman Vierkant, 325–50. Oxford: Oxford University Press.
- Vierkant, Tillman, Julian Kiverstein, and Andy Clark. 2013. "Decomposing the Will: Meeting the Zombie Challenge." In *Decomposing the Will*, edited by Andy Clark, Julian Kiverstein, and Tillman Vierkant, 1–30. Oxford: Oxford University Press.
- Waller, Bruce. 2011. *Against Moral Responsibility*. Cambridge, MA: MIT Press.
- Wegner, Daniel. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.
- Wolf, Susan. 1990. *Freedom Within Reason*. Oxford: Oxford University Press.
- Zahl, Simeon. 2020. *Holy Spirit and the Christian Experience*. Oxford: Oxford University Press.