

3 I Did That! Biomarkers of Volitional and Free Agency

Robyn Repko Waller and Maj. Allison Brager

Late at night, I opened a second bottle of wine. Did I do that? How can I tell whether I was the agent or whether my conscious choice caused me to drink more? What did cause me to drink more if not my conscious will? Was I too tired to think straight? This chapter attempts to answer questions such as these.

In section 3.1, we sketch a philosophical account of the significance and nature of agency, from both metaphysical and phenomenological points of view. In section 3.2, we discuss how the neuroscience of agency—how agents in fact exercise their agency—can bear upon philosophy of agency—accounts of what capacities or conditions would enable or prohibit free agency. Here, we review established and recent neuroscientific findings on agential control and the experience of agency. In section 3.3, we seek to establish a new avenue in interdisciplinary research on agency. Here, we outline how sleep deprivation can impact sense of agency (“I did that!”). In drawing this connection, we propose new research that bridges sleep research with well-established cognitive neuroscientific paradigms concerning voluntary action. Finally, we consider the importance of such new avenues for philosophical theories of free will and moral responsibility.

3.1 Metaphysics and Phenomenology of Agency

Sense of agency is an inextricable feature of human life. When we decide what to do—for instance, decide to pick up that cup of tea—we typically unreflectively believe that our decision or choice is up to us. When we move about in the world—for example, cycling along the roadside—we typically unreflectively believe that it is up to us when and how we move. Here, “up to us” indicates a kind of control over our decisions and actions. Intentional

decisions and actions are goal-directed activities over which the agent has some control.¹ We humans as social beings take it that, typically, other humans too have this kind of control over their decisions and actions. Moreover, we take it that humans extend this control not only over innocuous matters, such as sipping tea, but also over morally weighty ones, such as whether and how to care for and treat our dependents, affiliates, and distant others.

If we believe that at least some of our decisions and actions, especially our intentional bodily movements and their consequences, are up to us, we may believe that we are freely deciding or acting, at least at times. Accounts of free will differ widely, but most free will theorists take free will to be the ability to act freely. An agent who has free will has the appropriate kind of control over her decisions and actions to be an apt target of responsibility practices (see, e.g., Fischer & Ravizza, 1998). If one is an apt target of our moral responsibility practices, then one qualifies as the type of agent deserving of moral blame (for a morally bad action) and moral praise (for a morally good action; see, e.g., Fischer & Tognazzini, 2011; Watson, 1996). Here, by “responsibility practices,” we will understand these practices broadly to include not only moral practices of blaming or praising an agent, but also legal practices of punishment. Others take the control typically associated with responsibility to be valuable in its own right as grounding our human nature as self-governing agents.²

Exercising one’s free will involves control, or influence, over oneself. But what exactly must an agent control in order to exercise free will? One answer is that free will requires control over one’s *actions* (and perhaps the consequences of those actions). An agent exercises control over her action insofar as whether she does the action depends on her internal state of mind. This sense of free will is sometimes referred to as *free action*. Here, by “action,” we mean overt action, such as a bodily movement and the effects of that movement. This makes sense. When we hold people responsible—for example, blame them, punish them, feel gratitude toward them—we tend to hold them responsible for their conduct, or actions, and the repercussions of those actions. So, on (at least) a backward-looking notion of responsibility, we care about whether the agent had control over her actions. Why?

For instance, suppose on a crowded sidewalk Jan shoves into a group of pedestrians, causing an elderly pedestrian in the group to fall onto a crosswalk and sustain a serious injury. If we come to find out that Jan had

a seizure and her “shoving” was the result of that unexpected seizure, it would be strange and inappropriate to feel (or express) blame or anger or indignation at Jan. Likewise, we may take a similar view of the incident if we come to find out that Jan herself was pushed from behind by a crowd crush, thereby causing her to shove into the pedestrians. Although the incident was unfortunate for all involved, Jan’s bodily movement and the resulting shove were not her own or “up to her”—she, the agent, did not authorize or causally initiate the action, she did not desire or intend to shove into others or recklessly endanger them. She did not control her action.

Now, suppose instead that circumstances were different: Jan is late to a Broadway play, one for which she has paid significant money. She doesn’t want to miss the curtain’s going up and the doors closing for seating. She runs down the crowded sidewalk and, knowingly, straight into a group of pedestrians. The elderly pedestrian falls and sustains a serious injury. Here, of course, the group of pedestrians and especially the injured party should be permitted to, and likely would, express blame, anger, resentment (from the injured party), and indignation (from the spectators) at Jan for her reckless action and the resulting injury. To achieve her goal of not missing the opening curtain, Jan intentionally ran down the crowded sidewalk, knowing that she might well knock into others and that someone might get injured (even if Jan didn’t intend for others to get injured). Hence, her shoving of the pedestrians and the elderly pedestrian’s injury are the result of how she intentionally moved her body, an exercise of control over her actions. Moreover, the action expresses something about Jan as an agent: she was reckless, knowingly taking the risk of injuring others. In this way, the action was up to Jan.

So, it seems that the agent’s mental states—goals, desires, perceptions—at the time of action, in addition to how she moves her body, play a central role in whether and to what extent she controls her actions and so acts freely. But now one may reasonably ask the following: Even if an agent *acts freely*, how much control does the agent have over her state of mind? In particular, are agents’ decisions really up to them? That is, regardless of whether an agent performs a free action in the sense of her overt movements, we can still ask if the agent *decided freely*.³ We can make progress on these questions of control over one’s actions and decisions via the empirical work on the neuroscience of intentional bodily movement.

3.2 Neuroscience of Intentional Action

We noted that the kinds of actions that are “up to us,” so to speak, are those that come (at least partly) from the agent—and not wholly external sources—and further issue from the agent’s state of mind—her goals, desires, beliefs, and perceptions. There is a strong conceptual tie between actions (and consequences) that are up to us and actions for which agents are held responsible. For this reason, we don’t find it appropriate to blame Jan for injuring the elderly pedestrian in the event that Jan too was shoved from behind by the crowd. Her body moving as a result of a shove is a *passive and involuntary* movement caused by external forces. Furthermore, not all movements that are internally generated qualify as voluntary—think of Jan’s seizure-induced push into the crowd. Although the push into the crowd is the product of her (internal) neural and physiological processes, it isn’t produced in the right way—from her volitional states and capacities—to count as a voluntary action. It is only when we find out that Jan, the agent herself, initiated her shoving into the crowd that Jan is seen as having acted voluntarily and with the kind of control and mind-set that makes her worthy of blame for the injury that results. But how, then, do we sort out the neural mechanisms that underscore voluntary “up to us” movements as opposed to involuntary ones?

Fittingly, the neuroscience of agency provides a rich source of evidence concerning the mechanisms underlying voluntary actions, especially endogenously generated ones. Endogenously generated actions are actions for which the cue for acting does not come from the experimenter or other external stimuli but rather from within the participant—for example, within a designated window, a participant decides whether or when to execute an overt movement such as pressing a button. When we as agents execute endogenously generated movements, we experience ourselves as in control of those movements and their immediate effects in the world. We typically believe that it is indeed up to us when and how we act, at least some of the time. Moreover, we not only believe that what we do is the result of our decisions, but that what we do is frequently the result of decisions we are aware of. Hence, one commonsense assumption about endogenously generated actions, in line with widespread belief, is that the agent consciously “wills” or decides how and when to act, and that this decision initiates the process leading to action. Call the former phenomenology of control “sense of agency” and the latter assumption “the efficacy of conscious intentions.” In the next

section, we will discuss two broad neuroscientific paradigms that inform our understanding of the neural processes that subserve these aspects of agency. The intentional binding paradigm, detailed first here, provides an implicit measure of sense of agency. The Libet paradigm, outlined second, investigates the efficacy of conscious intentions in relation to the neural preparatory processes leading to action.

3.2.1 Intentional Binding as Implicit Sense of Agency

The most intuitive method in understanding sense of agency would be to solicit self-reports of the degree to which people see themselves instead of another as the agent who caused an event (e.g., a 100-point scale from “I allowed [event] to happen” to “I intended [event]”; Wegner & Wheatley, 1999). However, research on explicit sense of agency (“I did that”) has uncovered a cognitive bias impacting such judgments (Wegner & Wheatley, 1999). Participants who are asked in a social setting whether an outcome was the result of their agency or that of another person tend to misattribute unrelated events to their own agency, especially if the event is a positive one. Further, individuals exhibit reduced explicit sense of agency for negative outcomes than for positive ones (see, e.g., Bradley, 1978; Greenberg, Pyszczynski, & Solomon, 1992; Bandura, 1982, 2001).

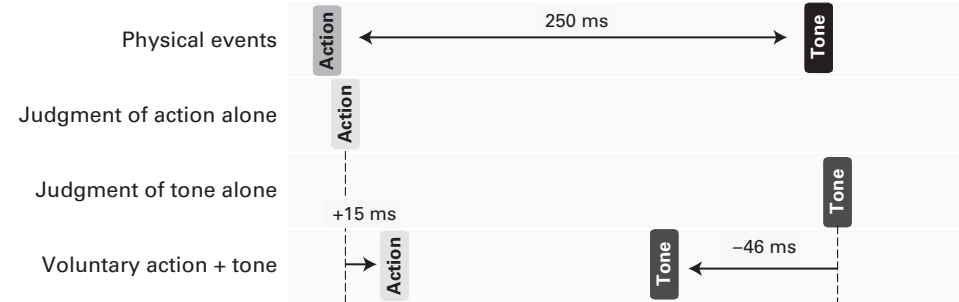
In contrast with explicit self-report, implicit measures of sense of agency, including intentional binding, attempt to avoid this self-serving bias (for a review of sense of agency, see Haggard, 2017). Intentional binding is understood in terms of how much the time interval between perceived time of action and perceived time of outcome (of that action) shifts in conditions of voluntary action (the active condition) compared to conditions of involuntary movement (the passive condition).

In the active experimental condition, participants are asked to execute an endogenously generated action, such as a button press, at a time of their choosing. A tone sounds 250 milliseconds after the button is pressed, and the actual time of their action is recorded. Participants watch a modified clockface as they complete the task. Participants report in separate blocks either the time of the button press or the time of the tone. In the passive experimental condition, using transcranial magnetic stimulation (TMS), experimenters induce a finger twitch in the participants with a tone sounding 250 milliseconds later. Similar to the active condition, participants report in separate blocks either the time of the finger twitch or the time of

the tone. In comparison, in baseline conditions, participants either execute an action (with no resultant tone) or listen for a tone (with no action) and report the time of the action or the tone. Intentional binding refers, then, to the phenomenon whereby the perceived time of action and tone shift closer to each other—bind together—when agents perform voluntary actions (relative to baseline time judgments; Haggard, Clark, & Kalogeras, 2002; Caspar et al., 2015; Haggard & Clark, 2003; see figure 3.1). Hence, intentional binding marks a difference in sense of agency between actions and non-actions, as well as among kinds of actions.

It should be noted that the designation of this temporal binding as action specific—indicative of an agent’s perception of her own voluntary action—is not univocal. In particular, there is an active debate as to the scope of intentional binding. Does the shift together in perceived timing of events accompany just actions and actional outcomes or, rather, more broadly the timing of physical causes with their sensory effects? In support of the latter wider scope, some studies are suggestive of temporal binding as a marker of observed physical causation more broadly (Buehner & Humphreys, 2009; Buehner, 2012, 2015; Kirsch, Kunde, & Herbort, 2019; Suzuki et al., 2019).

In favor of the narrower-scope view of intentional binding as a marker of perceived self-agency, several studies link intentional binding specifically to the motor system and to explicit judgment of agency. Such narrower-scope proponents argue that the kinds of experimental task features that enhance or attenuate intentional binding provide evidence for this action-specific domain of intentional binding (Bonicalzi & Haggard, 2019; Haggard, 2019). For instance, intentional binding is enhanced—that is, the perceived time of the action and the perceived time of the outcome are closer together—to the degree that participants have more courses of action available from which to choose. In one study, the intentional binding effect for voluntary actions was strongest when participants chose from seven button response options, and a three-button response choice evidenced enhanced intentional binding over the one-button option (e.g., push the prescribed response button when you feel like it; Barlas & Obhi, 2013). This finding of open versus instructed action was replicated and extended to support that not only the availability of alternative options but also the perceived valence of the option can impact intentional binding. Participants who are given a “free” choice of options evidence the strongest intentional binding effect when

Voluntary-action condition**Control condition: involuntary TMS-induced movement****Figure 3.1**

This figure illustrates the intentional binding effect. Participants watch a sped-up analog clock. In the experimental voluntary action condition, participants execute an endogenously generated movement, a button press, at a time of their choosing. A tone sounds 250 milliseconds after the button is pressed. Using the clock, in separate blocks, participants judge either the time of the action or the time of the tone. In the baseline conditions, participants execute a button press with no resultant tone or hear a tone with no preceding movement. Participants judge the time of the stand-alone event. In the involuntary TMS movement condition, experimenters induce a finger twitch in the participants with TMS, with a tone sounding 250 milliseconds later. Like the voluntary condition, in separate blocks, participants judge the time of the time of the twitch or the time of the tone. In voluntary active conditions, the time of the movement and the time of the tone bind together in perceptual temporal space compared to baseline and involuntary conditions.

the outcome of action is perceived as pleasant (here operationalized as the pleasantness of the tone). Further, the availability of alternatives and the valence of the outcome influence the degree to which participants explicitly feel that they are in control (Barlas, Hockley, & Obhi, 2018). The emotional valence, positive or negative, of an outcome vocal tone more generally affects intentional binding for voluntary actions, with enhanced intentional

binding for positive valence tones and attenuated intentional binding for negative tones (Yoshie & Haggard, 2013, 2017; but see Moreton, Callan, & Hughes, 2017, for conflicting findings). However, importantly, the valence of an outcome for a third-party only, and not the agent, does not impact the agent's implicit sense of agency in terms of intentional binding (Haggard, 2019).⁴ In line with these results, Borhani, Beck, and Haggard (2017) found that more painful (temperature) stimulus outcomes for the agent attenuated intentional binding, but this attenuation was weakened for free choice (trials with options) and active (vs. passive) trials. Hence, having options and the value of one's options seem to play a role in generating a stronger (implicit) sense of agency over one's actions and outcomes.

Interestingly, social and emotional factors that impact the likelihood of acting (Latane & Darley, 1968) and the exercise of self-control also modulate both explicit and implicit sense of agency. Recent work suggests that explicit judgments of control are impacted by the presence of others. Participants report reduced feelings of control over consequences of their actions when acting in the company of others who did not act but could have performed the action in place of the agent (Beyer et al., 2017). Implicit sense of agency in terms of intentional binding is attenuated in the lab in a number of situations that impact both exercised and experienced agential control in real life. Participants evidence a weakened temporal binding between voluntary actions and their outcomes when acting under orders in a coercive paradigm (Caspar et al., 2016), when induced to feel fearful or angry (Christensen et al., 2019), and when acting under the influence of alcohol (De Pirro et al., 2019).

We have surveyed how intentional binding can capture aspects of our experience of control and how this experience varies, but what is the neural basis of this experience? Brain studies demonstrate that the intentional binding effect is produced or affected by activity in the frontal and parietal cortex. Noninvasive continuous theta burst stimulation (cTBS) of the pre-supplementary motor area (pre-SMA) using TMS—a procedure that inhibits locally targeted neural activity—disrupts intentional binding compared to control cTBS of other areas (Moore et al., 2010). Furthermore, single-pulse TMS over the left inferior parietal lobe disrupts sense of agency for outcomes of voluntary actions when the participants choose among action options. Here, Chambon, Moore, and Haggard (2015) argue that this disruption to

implicit sense of agency in a choice paradigm may reflect the disruption of the functional network involving the dorsolateral prefrontal cortex (dlPFC) and the inferior parietal lobe subserving premovement action selection.

This research on sense of agency still does not give us a full picture of how an action can be up to the agent. As Bonicalzi and Haggard (2019) aptly highlight, self-initiated actions—such as the kind at issue in the free will and responsibility literature—are not characterized simply by the presence of subjective aspects of agency such as explicit and implicit sense of control. Rather, to appreciate fully how an action can be up to the agent, we will need grounding in the kinds of internal activity that are causal contributors and/or constitutive of agent-authored action. Accordingly, the next section will review the neuroscience of action initiation for voluntary (i.e., endogenously generated) actions.

3.2.2 Libet's Paradigm and the Initiation of Intentional Action

The commonsense model, as discussed above, takes it that the kinds of actions that we classify as up to us are not just those that are internally generated, but those that, in addition, are *agent initiated*. This claim in the metaphysics of free action suggests that for our movements to be up to us, the agent should be the causal initiator or at least a prominent causal contributor to the preparation for movement and subsequent movement. For self-initiated actions, the agent consciously wills or decides when and how to act, and that volition or decision initiates and (at least in part) causes the preparatory processes leading to movement. For instance, when I pick up my cup of coffee for a sip, it seems to me that first I, the agent, consciously decide to have a sip of coffee now, and then and only then do I “hand off” control to my neural action preparation circuits to bring it about that I pick up the coffee cup. The empirical question thus became: When does the conscious intention to move appear in relation to neural preparation to move and subsequent movement?

This question has been addressed by a series of experiments. Using electroencephalography (EEG), Kornhuber and Deecke (1965) discovered that when participants make simple movements in the lab, these movements are preceded by a slow negative buildup of precentral and parietal cortical activity bilaterally, termed the “Bereitschaftspotential” or “readiness potential” (RP). The RP is taken to be indicative of preparation for action and so a motor-specific neural signature. Libet, Wright, and Gleason (1982) extended these

findings to demonstrate that the RP preceded specifically endogenously generated simple movements (e.g., wrist or finger flexing). How and when does the agent's conscious intention to move come on the scene for such movements?

Libet and colleagues (1983) asked participants to perform a series of endogenously generated spontaneous (not preplanned) wrist flexes as they felt the desire or intention to do so.⁵ The beginning of each trial was signaled by a beep. Post flexing, participants reported the time of their awareness of their intention to flex in each trial by indicating where a moving spot was on a modified clockface when they were first aware of their intention to flex. This modified clock—now referred to as a Libet clock—makes a complete revolution every 2.56 seconds.⁶ Averaging over forty trials, the average time when the RP started to ramp up for these trials was about a half second (517 milliseconds) before recorded movement. In contrast, the average reported time of first awareness of intention to move, labeled “W,” occurred about 206 milliseconds prior to movement (“M”). Libet and colleagues found, then, that the RP preceded the reported time of conscious intention to move by around a third of a second (see figure 3.2).

Libet and co-authors (1983) and Libet (1985) took these findings to indicate that unconscious brain processes decide (form an intention) to flex, with conscious intentions to flex temporally later. On the basis of his results, some theorists—including Libet himself (e.g., 1999)—have concluded that our conscious intentions never causally initiate our overt actions (see also

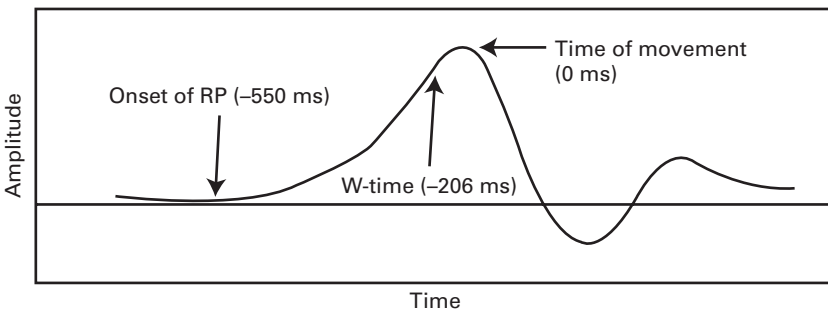


Figure 3.2

Libet's experimental results. The diagram shows the results of Libet and colleagues' (1983) experiment. The readiness potential (RP) onset occurs approximately 550 milliseconds prior to time of movement, labeled here as “0,” and approximately 300 milliseconds prior to the average reported time of first awareness of intention to act, labeled here as “W.”

Wegner, 2002). Libet meant this account to cover *all* overt actions, including complex movements. Notably, Libet did not test the latency of the RP and W time in the lab for complex actions. However, Libet reasoned that if an RP is present for all overt movements and W appears after the onset of the RP in his studies, then RPs occur prior to conscious intentions and initiate action for all intentional actions, even complex ones (Libet, 2005, p. 560). This result is, notably, in tension with the commonsense model of agency where the agent via a conscious intention initiates and/or causes the preparation for action.⁷

In the years following the study by Libet and colleagues (1983), Libet-style neuroscientific investigation of free agency has become a booming experimental enterprise. Others have replicated and extended the basic finding that brain activity associated with action preparation precedes W time (including, e.g., an fMRI-based version [Soon et al., 2008] and a single-cell recording version [Fried, Mukamel, & Kreiman, 2011]). To cover all of the careful work, empirically and philosophically, on the broadly construed Libet studies here would not be feasible.⁸ Nonetheless, we will briefly survey parts of this literature via an assessment of one pressing line of argument for free will skepticism relying on Libet-like studies—what we call the “*unconscious initiation threat to free agency*.”⁹

Recall the claims by Libet and others that conscious intentions do not initiate the preparatory process leading to action; rather, unconscious brain processes do.¹⁰ Call this the *empirical* premise. Supplement this empirical premise with the following prescriptive *metaphysical* premise: conscious intentions, not unconscious brain activity, must initiate actions for actions to be free.¹¹ From a commonsense perspective, it seems that my free agency is threatened if my unconscious brain always settles what I am to do before I am aware of my decision or consciously weigh in via deliberation.¹² That kind of internally generated action isn’t enough for the action to be up to me the agent. Put together, these two premises yield the conclusion that agents do not in fact act freely.

The empirical premise above is dependent on two substantive assumptions. First, the neural activity captured by the RP is indicative of neural preparation for a committed particular movement. Second, Libet’s interpretation of the findings rests on taking W, the averaged reported time of first awareness of intention to act, to be a reliable experimental measure of the time of a conscious intention to act. Otherwise, we cannot be sure that W occurring

after the onset of the RP in the experimental timeline yields any significance for the timing of the conscious intentions in producing action.¹³ Here, we will briefly review the evidence for and against these assumptions, and so for and against the unconscious initiation threat to free agency.

Let us first evaluate the claim that the RP is the neural signature of a (perhaps initially unconscious) practical intention. If the RP encodes an intention to move, then one might posit that there ought to be a relation between the RP qua intention to move and agents' awareness of an intention to move. Since participants in the Libet paradigm report the timing of their awareness of their intention to move, W time, one might expect a significant relationship between the time of onset, or initial ramp-up, of the RP and W . However, the evidence for covariance of RP and W is weak at best. While Jo and colleagues (2015) found that the RP co-varies with W time for a particular sub-population, committed meditators,¹⁴ others have failed to find a significant relationship between RP and W for general populations of agents (Haggard & Eimer, 1999; Schlegel et al., 2013; Alexander, Schlegel, Sinnott-Armstrong, Roskies, Tse, et al., 2015). Intriguingly, there is some evidence, albeit not univocal, that the timing of the lateralized RP (LRP) is correlated with W time (Schlegel et al., 2013; Alexander, Schlegel, Sinnott-Armstrong, Roskies, Tse, et al., 2015). When an agent makes a unilateral movement (e.g., left movement or right movement), the LRP represents the difference in electrical potential to the left and right of the vertex prior to that movement.

Even if the timing of the RP ramp-up and reported time of conscious intention are not correlated, one might still expect that RP is sufficient for the *presence* of a conscious intention to act. That is, if an agent exhibits an RP prior to voluntary movement and the RP is related to an intention to act, then the agent will have *some* conscious intention related to her so acting (at some time prior to movement). Even this assumption, taken for granted in earlier Libet-inspired work, has been thrown into question with new empirical evidence supporting similar RP and LRP profiles for both aware and hypnotized participants (Schlegel et al., 2015). Here, the hypnotized participants lack awareness of an intention to act but still show a RP ramp-up prior to action. Recent experimental redesigns and analysis, though, have revived the case for the RP representing intention to move. Using an online—premovement—measure, Parés-Pujolràs and colleagues (2019) argue for a significant relation between the RP and conscious intention to act. Parés-Pujolràs and colleagues (2019) found RP-like ramp-up prior to

the button press to report in-progress preparation to move.¹⁵ This on-line report of conscious intentions offers a promising new prospective measure of reported time of conscious intention, especially as recent work has supported that the traditional retrospective off-line measure, *W* time, is vulnerable to experimental task order effects and so may be worryingly unreliable (Dominik et al., 2017; Sanford et al., 2020). This work may renew worries that *W* is a good operationalization of awareness of intention to move. Moreover, the finding that externally triggered movements fail to produce an RP adds to the support for the RP's link to settledness of action, in particular endogenously generated action (Jahanshahi et al., 1995).

A weaker interpretation of the Libet results is that even if the RP is not the biomarker of intention to act, it may well still be the biomarker of preparation to act or a bias toward a particular action (Mele, 2009, 2012; Shepherd, 2015). That is, the RP does not signify commitment to an action plan, but rather general readiness or, alternatively, inclination to an action plan. Several studies have challenged this motor-domain interpretation of the RP, finding that the analogue of the RP is present when participants complete non-movement-based cued cognitive decision tasks (e.g., mental addition and subtraction, imagined movement; Alexander, Schlegel, Sinnott-Armstrong, Roskies, Tse, et al., 2015; Alexander, Schlegel, Sinnott-Armstrong, Roskies, Wheatley, et al., 2016; Raś et al., 2019). In contrast, the analogue of the LRP is present for only motor tasks. In support of a motor functional role, Travers and colleagues (2020) argue that while RP-like events were ubiquitous in the EEG recordings preceding both self-initiated and no movement tasks, such RP-like events are likely to be “false-positives” generated by distinct neural mechanisms than the RPs generated by voluntary motor preparation. Here, we might agree that regardless of its motor-specific or more general function, the RP is a neural signature of initiation of voluntary (i.e., endogenously generated) action where we can broaden the tent of voluntary as inclusive of not just overt bodily movement but also purely mental endogenously generated activity, such as mental math and imagining.¹⁶

Taken together, the threat of unconscious initiation of action from the Libet results seems underwhelming. We at best have good grounds for taking the RP to be neural preparation for voluntary actions. However, on that count, the Libet task appears to remain a good neuroscientific framework for understanding voluntary agency. In particular, the Libet paradigm is solidly established for investigating the neural underpinnings of action initiation

for simple self-authored actions. In this vein, in section 3.3, we will begin to sketch how the Libet paradigm can shed light on agency-related changes during conditions of sleep deprivation. Citing known results on agency and decision making during sleep reduction, we will construct a testable theory of the changes in the amplitude and latency of the RP prior to endogenously generated movement in sleep-deprived and non-sleep-deprived individuals. Further, relying on the above cited literature on intentional binding, we will piece together how sense of agency is impacted when agents are subject to reduced sleep. In this way, we aim to inspire new avenues of understanding on how sleep-deprived agents, a sizable contingent of the population, may exercise altered control of their actions and further experience their agency in a deviant fashion.

3.3 Effects of Sleep Deprivation on Reasoning and Agency

False confessions—suspects legally admitting to committing actions that they did not perform—is a popular trope of crime shows and sometimes, more depressingly, real-life criminal cases. Confessing to a crime is, in a sense, admitting ownership of the bodily actions and consequences that constitute the criminal act. Yes, I did that.

Are such false confessions more likely when suspects are sleep deprived? Frenda and colleagues (2016) kept participants awake for twenty-four hours. Those participants who self-reported high levels of sleepiness in the study were four-and-a-half times more likely than those who reported low levels of sleepiness to sign a confession that they had hit the escape key and thereby deleted the study data (when they had, in fact, not done so). Perhaps, though, as the researchers themselves note, these participants were not so much retrospectively reporting their genuine belief that they hit the escape key, but instead suffering from a reduced capacity to inhibit behavioral impulses and to resist suggestive influence.

Sleep deprivation might have similar effects on many other kinds of action ascription. We will discuss how sleep deprivation warps sense of agency in terms of intentional binding. We will then consider how sleep deprivation affects action control and action initiation, and how these effects bear upon action in a moral context. These two ways in which neurobiological regulation of responsible agency unravel with sleep deprivation are of prime importance for the sizable subpopulation of agents who suffer

from insufficient sleep yet must perform in high-stakes contexts, such as military personnel, shift workers, college students, psychiatric populations, parents of young children, and others.

3.3.1 Sleep Deprivation

Sleep deprivation is multifaceted. Here, we will focus on *acute sleep deprivation* and *chronic sleep restriction*. Acute sleep deprivation describes a situation or environment in which wakefulness extends over twenty-four hours (Goel, Basner, et al., 2013). Executive function is impaired in a time-dependent manner across twenty-four hours of forced wakefulness (Lo et al., 2012). Because some individuals show resiliency to twenty-four hours of total sleep deprivation (TSD), researchers have found that resiliency reaches a tipping point after forty-eight hours of TSD and beyond (McCauley et al., 2013). Under this paradigm, wakefulness has been extended for sixty-two hours on average (Goel, Basner, et al., 2013), but it has extended as far as ninety-six hours (Goel, Basner, et al., 2013).

Executive function is also severely impaired by chronic restricted sleep (Goel, Abe, et al., 2014). Chronic sleep restriction describes a situation or environment in which an individual is awake for more than eighty percent of a twenty-four-hour period across more than five days (Goel, Abe, et al., 2014). Five hours of sleep for seven days is enough to put executive function in a “danger zone” if subjects report that they need seven to eight hours of sleep a night in order to perform normally (measured by self-report).

In this section, we apply these sleep paradigms first to the Libet paradigm for initiation of action and then to the intentional binding paradigm for sense of agency. Finally, we look at applications of these sleep paradigms to situations in which the agent faces a morally significant decision with real-world relevance. Although empirical work connecting sleep paradigms and the two neuroscience agency paradigms has yet to be conducted, we survey related results and sketch what we might expect to find if we were to conduct the Libet and intentional binding paradigms with acute sleep-deprived and chronic sleep-restricted agents.

3.3.2 Sleep Deprivation and Agency: What We Know and Future Directions

Studies support compromised cognitive performance under insufficient sleep and also suggest that the functional neuroanatomy of this cognitive

disruption overlaps with the functional neuroanatomy of task competence for intentional binding and the Libet task. Specifically, these studies show a decline in performance of tasks requiring: (1) cognitive flexibility—deployment of mental resourcefulness in order to adapt to a dynamic environment (Honn et al., 2018); (2) flexible adaption—attentional control for “where” and “what” (Whitney et al., 2017); and (3) response inhibition (Zhao et al., 2019). In these studies, the dlPFC and pre-SMA were shown to underlie compromised functional connectivity manifest from insufficient sleep (Simon et al., 2017; Yang et al., 2018). As detailed above, both the dlPFC and pre-SMA are critical in order to perform tasks under the Libet and intentional binding paradigms. For instance, the early RP, the biomarker of initiation of agency in the Libet paradigm, is found bilaterally in the pre-SMA (Fried et al., 1991). Further, although sleep is thought to be a brain-centric process, this hypothesis has been challenged by the discovery of a sleep regulatory pathway in skeletal muscle (Ehlen, Brager, et al., 2017). This finding bears special relevance to the neuroscience of overt bodily actions, as considered in section 3.2, and suggests that the effects of sleep deprivation on action production go well beyond the cognitive pathways we review below.

How does sleep impact the neural activity of these areas to compromise cognitive performance? Slow-wave brain activity (SWA), measurable by EEG, is the basis for the recuperative value of sleep (Ehlen, Jefferson, et al., 2013). SWA include delta waves (0.5–4 Hz), characteristic of sleep and dream states, as well as theta waves (4–8 Hz), characteristic of states of drowsiness. In contrast, busy or active wakeful states are characterized by higher-frequency gamma (38–42 Hz) or beta waves (12–38 Hz), the former especially for instances of concentration (Abhang, Gawali, & Mehrotra, 2016). If a person is subject to restricted sleep, the temporal distribution of the EEG architecture changes such that SWA intrudes into the waking EEG (Vyazovskiy et al., 2011), in particular locally for the dlPFC and neighboring areas (Ehlen, Jefferson, et al., 2013). Specifically across extended wakefulness, the EEG is redistributed to include more theta compared to higher-frequency activity. Across recovery sleep following extended wakefulness, the EEG is redistributed to include more delta compared to higher-frequency activity (Ehlen, Jefferson, et al., 2013).

The Libet paradigm also relies on brain activity measurable via EEG. So, we could use it to measure how action initiation, understood as the RP, is likely to change under sleep deprivation. Combining the slow-wave

intrusion into extended wakefulness and the finding that the early RP is generated in the pre-SMA, bilaterally, we should expect that agents subject to acute sleep deprivation and perhaps chronic sleep restriction will have more positive (i.e., smaller amplitude) and later latency RPs than agents who are not sleep deprived. This pattern would indicate weaker and delayed intentions to act or preparation for action, depending on how the RP is interpreted. Further, we might also expect that agents who are sleep deprived on average have delayed *W* times, or reported times of first awareness of their intention to move, compared to non-sleep-deprived agents. However, this finding by itself would not necessarily indicate that awareness of one's intentions is impaired under lack of sleep. Whether this effect is specific to agency would depend on whether agents with insufficient sleep also evidence a delayed *M* time, reported time of their movement. If they do, then it is not the case that an aspect of sense of agency has been impacted by a lack of sleep as much as that more general attentional capacities are impaired with a lack of sleep. In any case, insofar as action control involves initiation of action and preparation for action, there are good grounds for thinking that sleep deprivation will reduce action control, at both the subpersonal and personal levels.

To answer the question of how sleep deprivation impacts sense of agency specifically, we can turn to how agents' performance on the intentional binding task might be affected by insufficient sleep. Previous research found that when participants were subject to twenty-four hours of forced wakefulness, participants gave enhanced explicit sense of agency judgments for others' actions (Hon & Po, 2016). But given the self-serving bias moderating explicit sense of agency judgments, it would be useful to know how sleep deprivation impacts implicit sense of agency. This issue can be considered with respect to both the action-binding and effect-binding elements of intentional binding under conditions of sleep deprivation (or sufficient sleep). *Action-binding* refers to the perceived shift forward of the time of action toward the outcome in voluntary conditions. *Effect-binding* refers to the perceived shift backward of the time of the outcome toward the action (cause) in voluntary conditions. We predicted above that agents who are acutely sleep deprived or chronically sleep restricted should on average have more positive RPs (with smaller amplitude in a negative direction) than non-sleep-deprived agents. Interestingly, more positive RPs are associated with attenuated intentional binding, particularly for effect binding (Jo et al.,

2014). Hence, we expect that insufficient sleep should lead to attenuated intentional binding, especially for effect binding. If so, agents who act under sleep deprivation would evidence weakened (implicit) sense of agency.

This weakened sense of agency at the level of perception of event timing may well be significant for morally weighty exercises of agency, as the phenomenology of agency—the experience of agency—is taken to ground the capacity for choice (cf. the debate on deliberative compatibilism; e.g., Nelkin, 2004) and the capacity for free decisions and actions. For instance, the seeming that one is deciding in the face of genuinely open alternatives and is in control of actions and outcomes drives the motivation and development of accounts of free will (McKenna & Coates, 2016). Changes in perception of one's agency, then, may impact the kinds of decisions and actions that agents attribute to themselves and hence whether observers attribute responsibility for those decisions and actions to those agents. Indeed, given that certain subpopulations are subject to insufficient sleep non-culpably as part of the conditions of their employment, we ought to pay attention to how sleep deprivation can impact agency and responsibility judgments in moral contexts and also what mediates these relationships. The latter question of mediation is beyond the scope of this review but may be relevant for how to fit together changes in agency and sense of agency with changes in moral judgment under sleep deprivation. In fact, here, there is a growing neuroscience literature on sleep deprivation and moral reasoning and judgment to which we now turn to conclude our review.

3.3.3 Sleep Deprivation and Moral Reasoning: What We Know and Future Directions

Though limited, a handful of studies to date have observed impairment in moral reasoning under sleep deprivation in healthy civilians (Killgore, Killgore, et al., 2007; Tempesta et al., 2012) and military officers (Olsen, 2013; Olsen, Palleson, & Eid, 2010). These studies did not explore the neural basis of their behavioral findings. However, similarities between the brain areas that are affected by sleep deprivation and those that are used for moral judgments suggest that sleep deprivation could have profound effects on moral judgment.

In the most recent decade, substantial salience has been placed on the study of PFC processes and their role in impaired performance under sleep deprivation. A simple keyword search for “sleep deprivation AND

prefrontal cortex AND humans” in the National Library of Medicine database (PubMed) revealed 169 articles. The negative consequences of sleep deprivation on PFC processes that have been discovered in the last year are multidimensional and include decreased oxygenation (Borragán et al., 2019), altered small-molecule metabolites (Bourdon et al., 2018), and decreased functional connectivity (Yang et al., 2018). Among subcortical regions, the nucleus accumbens (NAc) is also highly recruited and activated under conditions of sleep deprivation, making an evidence-based argument for sleep deprivation as a reward-seeking state (Volkow et al., 2012).

Similar processes and areas have been found to be important in moral judgment.¹⁷ Regions of the frontal cortex that show altered activity while considering moral dilemmas include the ventromedial PFC and the superomedial prefrontal cortex (from de Oliveira-Souza, Zahn, & Moll, 2015). Among subcortical regions, the NAc shows altered functional connectivity in response to moral dilemmas (de Oliveira-Souza et al., 2015). Lesion and psychopathology studies have identified the NAc as a key modulator of reward-seeking behavior (de Oliveira-Souza et al., 2015). Convergence of these separate studies thus lends credence to the hypothesis that decreased moral reasoning under sleep-deprived states is neurobiologically manifest from hyperactivation of rapid reward-seeking states.

Advanced neuroimaging techniques have also shown direct linkage of prefrontal and anterior subcortical regions with the subcortical limbic system, including the amygdala and hippocampus, and the surrounding temporal cortex, including the superior temporal sulcus (de Oliveira-Souza et al., 2015). These medial regions are altered by moral reasoning (from de Oliveira-Souza et al., 2015) and also sleep deprivation, as shown in independent studies (Goldstein et al., 2013).

Another brain region worthy of attention is the septum. Lesion studies in animal models have found a phenomenon known as “septal rage” (Caplan, 1973) as well as dysregulation of sleep processes (Gerashchenko, Salin-Pascual, & Shiromani, 2001), but parallel studies in humans are limited. Nevertheless, the identification of the septum as a modulator of moral behavior (Gorman & Cummings, 1992) suggests that emotional states related to the septum are involved in moral judgment and decision making. This connection lends further credence to the hypothesis that sleep deprivation can unravel moral judgment and decision making along with related emotions.

Additional evidence comes from neurochemical studies. Sleep–wake transitions are largely mediated by acetylcholinergic, monoaminergic, and GABA-ergic pathways (reviewed in Torterolo, Monti, & Vanini, 2019). Experiments have also demonstrated the impact of altered neurochemical tone on moral reasoning processes. The greatest emphasis has been placed on alteration of neurochemical and neuroendocrine tone through psychotropic medications that can dually serve as sleep- or wake-promoting medications, including: (1) serotonergic tone through citalopram, (2) norepinephrine tone through propranolol, (3) dopaminergic tone through L-DOPA and modafinil, (4) intranasal delivery of oxytocin, and (5) oral delivery of testosterone (Crockett, Clark, Hauser, et al., 2010; Crockett, Clark, Lieberman, et al., 2010; Terbeck et al., 2013; Levy et al., 2014; De Dreu et al., 2011; Montoya et al., 2013). The indirect consequences of these existing dual psychotropic and sleep–wake medications are a foundation for future research efforts.

Several of these neurotransmitters and brain areas are related to affect or emotion. So, that might be one key to understanding the relation between sleep and moral judgment. Studies of acute sleep deprivation of more than forty-eight hours in healthy individuals have found modest yet subclinical deviations in atypical emotional or affective states (Kahn-Greene, Killgore, et al., 2007) linked to moral reasoning, including: (1) reductions in perceived emotional intelligence (Killgore, Lipizzi, et al., 2008) and reactivity to moral dilemmas (Zohar et al., 2005); (2) elevations in interpersonal frustration (Kahn-Greene, Lipizzi, et al., 2006); and (3) poor moral judgment (Killgore, Killgore, et al., 2007). In all of these studies of acute sleep deprivation, these affective deviations in healthy individuals subside with adequate sleep during the recovery period, emphasizing that sleep alone is necessary and sufficient to stabilize affective states linked to moral reasoning.

Much more remains to be discovered. However, these studies at least provide tantalizing hints of new directions for research on how sleep deprivation can alter neurochemicals, neuronal activity, and brain networks that are essential for moral judgments, decisions, and emotions.

3.4 Concluding Remarks

In this chapter, we have argued that action initiation and taking oneself to be the author of events in the world are two hallmarks of what it is to be

an agent exercising control over one's actions and making one eligible for morally reactive attitudes. We then reviewed two well-established neuroscientific frameworks for the investigation of action initiation and sense of agency: the intentional binding paradigm and the Libet paradigm. We next outlined how the subpopulation of agents who act under sleep deprivation might perform in those paradigms via an assessment of the current state of the literature on cognitive performance and the functional neuroanatomy of sleep deprivation. Finally, we reviewed the extant work on sleep deprivation and moral reasoning, with an eye toward how otherwise neurotypical agents who suffer from insufficient sleep might alter their moral reasoning and moral behavior in extended wakefulness.

We believe that such a review fills a lacuna in research and that future work connecting the neuroscience of agency paradigms with sleep paradigms is imperative. Several subpopulations suffer from insufficient sleep, including military personnel, shift workers, college students, doctors, and others in high-stress occupations. Where insufficient sleep is a necessary aspect of the job, these individuals are not culpable for the cognitive changes induced by sleep deprivation. Suppose, as is hypothesized, that weakened initiation of agency and reduced sense of agency impacts moral judgment and morally significant action. In the philosophical literature on moral responsibility, other stressors are taken as affecting control of practical decision making and actions and so are excusing or exempting conditions for moral and legal responsibility (e.g., extreme emotion, certain psychiatric conditions; see P. F. Strawson, 1974, for the inspiration for work on excusing and exempting conditions). Hence, insofar as one is not culpable for one's sleep deprivation, ought we to mitigate blame for morally bad actions under sleep deprivation? This question is largely unexplored theoretically. The synthesized research here promises to spark a new strand of critical debate, both in philosophy and in applied contexts such as criminal law, as to the control and responsibility of agents for their actions under sleep deprivation. It could also motivate the development of more effective interventions for these populations who must endure sleep deprivation.

Notes

1. Nonhuman animals engage in goal-directed behaviors in some sense, but here the focus is a subset of goal-directed behaviors, human intentional actions.

2. So, too, this kind of control is distinctive of human nature insofar as we do not hold other species to be morally responsible for what they do.
3. One prominent example to illustrate the difference is Locke's (1975) locked room case. I am sitting in a room chatting with a friend. I prefer to stay chatting longer. Unbeknownst to me, the door to leave is locked. While I can freely decide to stay (and can freely decide to leave), I am not free to leave.
4. Here, we may worry that, given these results, to some degree, intentional binding as a marker of sense of agency also falls prey to a self-serving bias effect.
5. Libet and colleagues (1983) took this item of interest, discussed here as the conscious intention to flex, to be the broader disjunction of intention, urge, or desire to flex.
6. The Libet clock is same time perception measure clock adopted by the intentional binding paradigm, as referenced in II.1.
7. Libet did posit a more restricted role for conscious intentions in the action production process. In his view, although conscious intentions do not initiate neural preparation for movement, an agent may consciously veto that neural preparation prior to movement. That is, the agent can consciously decide not to move and so not move (Libet, 1999, 2005).
8. For a recent wide-ranging review of the neuroscience of free will based on Libet's studies, see Waller (2019). For a book anthology of empirical and philosophical work on the Libet paradigm, see Sinnott-Armstrong and Nadel (2011). For a book-length treatment of science and free will, see Mele (2009); in popular writing format, Mele (2014a); and in dialogue form, Mele (2014b).
9. We will not discuss here what one of the present authors (and others) have labeled elsewhere the dualist threat or the epiphenomenalism threat (e.g., Mele, 2014c; Nahmias, 2010; Waller, forthcoming). These are interesting and important issues but do not directly relate to the subpopulation of agents that are the focal point of this chapter in section 3.3, sleep-deprived agents.
10. This claim is independent from the claim that conscious intentions do not causally contribute to action production. It is consistent with the unconscious initiation threat that conscious intentions are causally efficacious downstream in either producing or vetoing bodily movement post unconscious preparation for movement. To deny this further causal role of conscious intentions in action production on the basis of the Libet results is to embrace the epiphenomenalism threat (see, e.g., Waller, 2019).
11. To be more precise, for those actions to be free in the basic sense. For an action to be a basically free one, it does not derive its free status from other earlier basically free actions. Hence, it is open to those who endorse this metaphysical claim to also

hold the following: if conscious mental states must be among the causes of a basically free decision, then other unconsciously produced decisions and actions can be free in virtue of their relation to those basically free decisions.

12. As Mele (2009) highlights, there is empirical work by Gollwitzer and colleagues supporting the causal efficacy of conscious distal intentions. Distal intentions are intentions to do something later—as opposed to intentions to do something now. Participants who form intentions on how and when to accomplish a goal later are significantly more likely to follow through on those plans (Gollwitzer, 1996; Gollwitzer & Brandstätter, 1997; Gollwitzer, 1999).

13. For a recent discussion of whether W , an average of reported times, accurately reflects the actual onset of a conscious intention to act now, see Mele (2010a, 2010b, 2010c, 2012). Herdova (2016) examines whether such experimental setups require the participant to have any intentions to act now at all.

14. Committed meditators—the experimental group in Jo, Hinterberger, and colleagues (2015)—were those who had at least three years of experience practicing mindfulness meditation.

15. See also Schultze-Kraft and colleagues (2020) for another innovative cueing approach that supports a relationship between the RP and presence of a conscious intention to move.

16. To take the RP as indicative of noncommittal preparation or bias toward voluntary action still conflicts with those who argue that the RP represents passive, not active (i.e., not true preparatory), activity. In particular, Schurger, Sitt, and Dehaene (2012) have pushed this skepticism about the active nature of the RP further, proposing the integration-to-bound model of RP. This model holds that, roughly, random walk type fluctuation leads to neural activity crossing the threshold for motor response (i.e., movement). The jury is still out on whether the integration-to-bound model accurately deflates the active, voluntary nature of the RP. For instance, Khalighinejad and colleagues (2018) found that the integration-to-bound model fits the recorded EEG signature of neural activity during endogenously generated movements *only if* an additional parameter corresponding to attention to task demands and cognitive effort—active neural noise—is added to the model.

17. See chapter 1 in this volume.

References

- Abhang, P. A., Gawali, B. W., & Mehrotra, S. C. (2016). *Introduction to EEG-and speech-based emotion recognition*. Cambridge, MA: Academic Press.
- Alexander, P., Schlegel, A., Sinnott-Armstrong, W., Roskies, A., Tse, P. U., & Wheatley, T. (2015). Dissecting the readiness potential: An investigation of the relationship

between readiness potentials, conscious willing, and action. In A. Mele (Ed.), *Surrounding free will: Philosophy, psychology, neuroscience* (pp. 203–233). New York: Oxford University Press.

Alexander, P., Schlegel, A., Sinnott-Armstrong, W., Roskies, A., Wheatley, T., & Tse, P. (2016). Readiness potentials driven by non-motoric processes. *Consciousness and Cognition, 39*, 38–47.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1–26.

Barlas, Z., Hockley, W. E., & Obhi, S. S. (2018). Effects of free choice and outcome valence on the sense of agency: Evidence from measures of intentional binding and feelings of control. *Experimental Brain Research, 236*(1), 129–139.

Barlas, Z., & Obhi, S. (2013). Freedom, choice, and the sense of agency. *Frontiers in Human Neuroscience, 7*, 514.

Beyer, F., Sidarus, N., Bonicalzi, S., & Haggard, P. (2017). Beyond self-serving bias: Diffusion of responsibility reduces sense of agency and outcome monitoring. *Social Cognitive and Affective Neuroscience, 12*(1), 138–145.

Bonicalzi, S., & Haggard, P. (2019). From freedom from to freedom to: New perspectives on intentional action. *Frontiers in Psychology, 10*, 1193.

Borhani, K., Beck, B., & Haggard, P. (2017). Choosing, doing, and controlling: Implicit sense of agency over somatosensory events. *Psychological Science, 28*(7), 882–893.

Borragán, G., Guerrero-Mosquera, C., Guillaume, C., Slama, H., & Peigneux, P. (2019). Decreased prefrontal connectivity parallels cognitive fatigue-related performance decline after sleep deprivation. An optical imaging study. *Biological Psychology, 144*, 115–124.

Bourdon, A. K., Spano, G. M., Marshall, W., Bellesi, M., Tononi, G., Serra, P. A., . . . & Cirelli, C. (2018). Metabolomic analysis of mouse prefrontal cortex reveals upregulated analytes during wakefulness compared to sleep. *Scientific Reports, 8*(1), 1–17.

Bradley, G. W. (1978). Self-serving biases in the attribution process: A re-examination of the fact or fiction question. *Journal of Personality and Social Psychology, 36*(1), 56–71.

Buehner, M. J. (2012). Understanding the past, predicting the future: Causation, not intentional action, is the root of temporal binding. *Psychological Science, 23*(12), 1490–1497.

Buehner, M. J. (2015). Awareness of voluntary and involuntary causal actions and their outcomes. *Psychology of Consciousness: Theory, Research, and Practice, 2*(3), 237.

Buehner, M. J., & Humphreys, G. R. (2009). Causal binding of actions to their effects. *Psychological Science, 20*(10), 1221–1228.

Caplan, M. (1973). An analysis of the effects of septal lesions on negatively reinforced behavior. *Behavioral Biology, 9*(2), 129–167.

Caspar, E. A., Christensen, J. F., Cleeremans, A., & Haggard, P. (2016). Coercion changes the sense of agency in the human brain. *Current Biology, 26*(5), 585–592.

Caspar, E. A., Cleeremans, A., & Haggard, P. (2015). The relationship between human agency and embodiment. *Consciousness and Cognition, 33*, 226–236.

Chambon, V., Moore, J. W., & Haggard, P. (2015). TMS stimulation over the inferior parietal cortex disrupts prospective sense of agency. *Brain Structure and Function, 220*(6), 3627–3639.

Christensen, J. F., Di Costa, S., Beck, B., & Haggard, P. (2019). I just lost it! Fear and anger reduce the sense of agency: A study using intentional binding. *Experimental Brain Research, 237*(5), 1205–1212.

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences of the United States of America, 107*(40), 17433–17438.

Crockett, M. J., Clark, L., Lieberman, M. D., Tabibnia, G., & Robbins, T. W. (2010). Impulsive choice and altruistic punishment are correlated and increase in tandem with serotonin depletion. *Emotion, 10*(6), 855.

De Dreu, C. K., Greer, L. L., Van Kleef, G. A., Shalvi, S., & Handgraaf, M. J. (2011). Oxytocin promotes human ethnocentrism. *Proceedings of the National Academy of Sciences of the United States of America, 108*(4), 1262–1266.

de Oliveira-Souza, R., Zahn, R., & Moll, J. (2015). 11 neural correlates of human morality: An overview. In J. Decety & T. Wheatley (Eds.), *The moral brain: A multidisciplinary perspective* (pp. 183–196). Cambridge, MA: MIT Press.

De Pirro, S., Lush, P., Parkinson, J., Duka, T., Critchley, H. D., & Badiani, A. (2019). Effect of alcohol on the sense of agency in healthy humans. *Addiction Biology, 25*(4), e12796.

Dominik, T., Dostál, D., Zielina, M., Šmahaj, J., Sedláčková, Z., & Procházka, R. (2017). Libet's experiment: Questioning the validity of measuring the urge to move. *Consciousness and Cognition, 49*, 255–263.

Ehlen, J. C., Brager, A. J., Baggs, J., Pinckney, L., Gray, C. L., DeBruyne, J. P., . . . & Paul, K. N. (2017). Bmal1 function in skeletal muscle regulates sleep. *eLife, 6*, e26557.

Ehlen, J. C., Jefferson, F., Brager, A. J., Benveniste, M., & Paul, K. N. (2013). Period-amplitude analysis reveals wake-dependent changes in the electroencephalogram during sleep deprivation. *Sleep, 36*(11), 1723–1735.

- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.
- Fischer, J. M., & Tognazzini, N. (2011). The physiognomy of responsibility. *Philosophy and Phenomenological Research*, *82*(2), 381–417.
- Frenda, S. J., Berkowitz, S. R., Loftus, E. F., & Fenn, K. M. (2016). Sleep deprivation and false confessions. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(8), 2047–2050.
- Fried, I., Katz, A., McCarthy, G., Sass, K. J., Williamson, P., Spencer, S. S., & Spencer, D. D. (1991). Functional organization of human supplementary motor cortex studied by electrical stimulation. *Journal of Neuroscience*, *11*(11), 3656–3666.
- Fried, I., Mukamel, R., & Kreiman, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron*, *69*(3), 548–562.
- Gerashchenko, D., Salin-Pascual, R., & Shiromani, P. J. (2001). Effects of hypocretin-1 injections into the medial septum on sleep and hippocampal theta. *Brain Research*, *913*(1), 106–115.
- Goel, N., Abe, T., Braun, M. E., & Dinges, D. F. (2014). Cognitive workload and sleep restriction interact to influence sleep homeostatic responses. *Sleep*, *37*(11), 1745–1756.
- Goel, N., Basner, M., Rao, H., & Dinges, D. F. (2013). Circadian rhythms, sleep deprivation, and human performance. *Progress in Molecular Biology and Translational Science*, *119*, 155–190.
- Goldstein, A. N., Greer, S. M., Saletin, J. M., Harvey, A. G., Nitschke, J. B., & Walker, M. P. (2013). Tired and apprehensive: Anxiety amplifies the impact of sleep loss on aversive brain anticipation. *Journal of Neuroscience*, *33*(26), 10607–10615.
- Gollwitzer, P. M. (1996). The volitional benefits of planning. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 287–312). New York: Guilford Press.
- Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*, 493–503.
- Gollwitzer, P. M., & Brandstätter, V. (1997). Implementation intentions and effective goal pursuit. *Journal of Personality and Social Psychology*, *73*, 186–199.
- Gorman, D. G., & Cummings, J. L. (1992). Hypersexuality following septal injury. *Archives of Neurology*, *49*(3), 308–310.
- Greenberg, J., Pyszczynski, T., & Solomon, S. (1982). The self-serving attributional bias: Beyond self-presentation. *Journal of Experimental Social Psychology*, *18*(1), 56–67.
- Haggard, P. (2017). Sense of agency in the human brain. *Nature Reviews Neuroscience*, *18*, 197–208.

- Haggard, P. (2019). The neurocognitive bases of human volition. *Annual Review of Psychology, 70*, 9–28.
- Haggard, P., & Clark, S. (2003). Intentional action: Conscious experience and neural prediction. *Consciousness and Cognition, 12*, 695–707.
- Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience, 5*, 282–385.
- Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research, 126*(1), 128–133.
- Herdova, M. (2016). Are intentions in tension with timing experiments? *Philosophical Studies, 173*(3), 573–587.
- Hon, N., & Poh, J. H. (2016). Sleep deprivation produces feelings of vicarious agency. *Consciousness and Cognition, 40*, 86–92.
- Honn, K., Hinson, J. M., Whitney, P., & Van Dongen, H. (2018). Cognitive flexibility: A distinct element of performance impairment due to sleep deprivation. *Accident Analysis and Prevention, 18*, 30070–30078.
- Jahanshahi, M., Jenkins, I. H., Brown, R. G., Marsden, C. D., Passingham, R. E., & Brooks, D. J. (1995). Self-initiated versus externally triggered movements: I. An investigation using measurement of regional cerebral blood flow with PET and movement-related potentials in normal and Parkinson's disease subjects. *Brain, 118*(4), 913–933.
- Jo, H.-G., Hinterberger, T., Wittmann, M., & Schmidt, S. (2015). Do meditators have higher awareness of their intentions to act? *Cortex, 65*, 149–158.
- Jo, H.-G., Wittman, M., Hinterberger, T., & Schmidt, S. (2014). The readiness potential reflects intentional binding. *Frontiers in Human Neuroscience, 8*, 421.
- Kahn-Greene, E. T., Killgore, D. B., Kamimori, G. H., Balkin, T. J., & Killgore, W. D. (2007). The effects of sleep deprivation on symptoms of psychopathology in healthy adults. *Sleep Medicine, 8*(3), 215–221.
- Kahn-Greene, E. T., Lipizzi, E. L., Conrad, A. K., Kamimori, G. H., & Killgore, W. D. (2006). Sleep deprivation adversely affects interpersonal responses to frustration. *Personality and Individual Differences, 41*(8), 1433–1443.
- Khalighinejad, N., Schurger, A., Desantis, A., Zmigrod, L., & Haggard, P. (2018). Precursor processes of human self-initiated action. *NeuroImage, 165*, 35–47.
- Killgore, W. D., Killgore, D. B., Day, L. M., Li, C., Kamimori, G. H., & Balkin, T. J. (2007). The effects of 53 hours of sleep deprivation on moral judgment. *Sleep, 30*(3), 345–352.
- Killgore, W. D., Lipizzi, E. L., Killgore, D. B., Rupp, T. L., Kamimori, G. H., & Balkin, T. J. (2008). Emotional intelligence predicts declines in emotion-based decision-making following sleep deprivation. *Sleep, 31*, A134.

- Kirsch, W., Kunde, W., & Herbolt, O. (2019). Intentional binding is unrelated to action intention. *Journal of Experimental Psychology: Human Perception and Performance*, 45(3), 378.
- Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, 284(1), 1–17.
- Latane, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology*, 10(3), 215–221.
- Levy, N., Douglas, T., Kahane, G., Terbeck, S., Cowen, P. J., Hewstone, M., & Savulescu, J. (2014). Are you morally modified?: The moral effects of widely used pharmaceuticals. *Philosophy, Psychiatry, and Psychology*, 21(2), 111–125.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8, 529–566.
- Libet, B. (1999). Do we have free will? *Journal of Consciousness Studies*, 6(8–9), 47–57.
- Libet, B. (2005). Do we have free will? In R. Kane (Ed.), *The Oxford handbook of free will* (pp. 551–564). Oxford: Oxford University Press.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain*, 106(3), 623–642.
- Libet, B., Wright, E. W., & Gleason, C. A. (1982). Readiness potentials preceding unrestricted spontaneous pre-planned voluntary acts. *Electroencephalography and Clinical Neurophysiology*, 54, 322–335.
- Lo, J. C., Groeger, J. A., Santhi, N., Arbon, E. L., Lazar, A. S., Hasan, S., . . . Dijk, D.-J. (2012). Effects of partial and acute total sleep deprivation on performance across cognitive domains, individuals and circadian phase. *PLoS One*, 7(9), e45987.
- Locke, J. (1975). *An essay concerning human understanding*. (P. Nidditch, Ed.) Oxford: Oxford University Press. (Original work published 1689)
- McCauley, P., Kalachev, L. V., Mollicone, D. J., Banks, S., Dinges, D. F., & Van Dongen, H. P. (2013). Dynamic circadian modulation in a biomathematical model for the effects of sleep and sleep loss on waking neurobehavioral performance. *Sleep*, 36, 1987–1997.
- McKenna, M., & Coates, J. D. (2016). Compatibilism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 ed.). Retrieved from <https://plato.stanford.edu/archives/win2016/entries/compatibilism/>
- Mele, A. R. (2009). *Effective intentions: The power of conscious will*. Oxford: Oxford University Press.

Mele, A. R. (2010a). Conscious deciding and the science of free will. In R. F. Baumeister, A. Mele, & K. Vohs (Eds.), *Free will and consciousness: How might they work* (pp. 43–65). Oxford: Oxford University Press.

Mele, A. R. (2010b). Libet on free will: Readiness potentials, decisions, and awareness. In W. Sinnott-Armstrong & L. Nadel (Eds.), *Conscious will and responsibility: A tribute to Benjamin Libet* (pp. 23–33). Oxford: Oxford University Press.

Mele, A. R. (2010c). Conscious intentions. In J. K. Campbell, M. O'Rourke, & H. Silverstein (Eds.), *Action, ethics, and responsibility* (pp. 85–108). Cambridge, MA: MIT Press.

Mele, A. R. (2012). Free will and neuroscience: Revisiting Libet's studies. In A. Suarez & P. Adams (Eds.), *Is science compatible with free will?*. New York: Springer.

Mele, A. R. (2014a). *Free: Why science hasn't disproved free will*. New York: Oxford University Press.

Mele, A. R. (2014b). *A dialogue on free will and science*. New York: Oxford University Press.

Mele, A. R. (2014c). Free will and substance dualism: The real scientific threat to free will? In W. Sinnott-Armstrong (Ed.), *Moral psychology: Free will and responsibility* (Vol. 4, pp. 195–207). Cambridge, MA: MIT Press.

Montoya, E. R., Terburg, D., Bos, P. A., Will, G. J., Buskens, V., Raub, W., & van Honk, J. (2013). Testosterone administration modulates moral judgments depending on second-to-fourth digit ratio. *Psychoneuroendocrinology*, *38*(8), 1362–1369.

Moore, J. W., Ruge, D., Wenke, D., Rothwell, J., & Haggard, P. (2010). Disrupting the experience of control in the human brain: Pre-supplementary motor area contributes to the sense of agency. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1693), 2503–2509.

Moreton, J., Callan, M. J., & Hughes, G. (2017). How much does emotional valence of action outcomes affect temporal binding? *Consciousness and Cognition*, *49*, 25–34.

Nahmias, E. (2010). Scientific challenges to free will. In C. Sandis & T. O'Connor (Eds.), *A companion to the philosophy of action* (pp. 345–356). Chichester, UK: Wiley-Blackwell.

Nelkin, D. K. (2004). Deliberative alternatives. *Philosophical Topics*, *32*(1/2), 215–240.

Olsen, O. K. (2013). The impact of partial sleep deprivation on military naval officers' ability to anticipate moral and tactical problems in a simulated maritime combat operation. *International Maritime Health*, *64*(2), 61–65.

Olsen, O. K., Palleson, S., & Eid, J. (2010). Impact of partial sleep deprivation on moral reasoning in military officers. *Sleep*, *33*, 1086–1090.

Parés-Pujolràs, E., Kim, Y. W., Im, C. H., & Haggard, P. (2019). Latent awareness: Early conscious access to motor preparation processes is linked to the readiness potential. *NeuroImage*, *202*, 116140.

Raś, M., Nowik, A. M., Klawiter, A., & Króliczak, G. (2019). When is the brain ready for mental actions? Readiness potential for mental calculations. *Acta Neurobiologia Experimentalis*, *79*, 386–398.

Sanford, P., Lawson, A. L., King, A. N., & Major, M. (2020). Libet's intention reports are invalid: A replication of Dominik et al. (2017). *Consciousness and Cognition*, *77*, 102836.

Schlegel, A., Alexander, P., Sinnott-Armstrong, W., Roskies, A. Tse, P., & Wheatley, T. (2013). Barking up the wrong tree: Readiness potentials reflect processes independent of conscious will. *Experimental Brain Research*, *229*(3), 329–335.

Schlegel, A., Alexander, P., Sinnott-Armstrong, W., Roskies, A. Tse, P., & Wheatley T. (2015). Hypnotizing Libet: Readiness potentials with non-conscious volition. *Consciousness and Cognition*, *33*, 196–203.

Schultze-Kraft, M., Parés-Pujolràs, E., Matic, K., Haggard, P., & Haynes, J. D. (2020). Preparation and execution of voluntary action both contribute to awareness of intention. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1923), 20192928.

Schurger, A., Sitt, J. D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(42), E2904–E2913.

Shepherd, J. (2015). Scientific challenges to free will and moral responsibility. *Philosophy Compass*, *10*(3), 197–207.

Simon, E. B., Lahav, N., Shamir, R., Hendler, T., & Maron-Katz, A. (2017). Tired and misconnected: A breakdown of brain modularity following sleep deprivation. *Biological Psychiatry*, *81*(10), S322.

Sinnott-Armstrong, W., & Nadel, L. (Eds.). (2011). *Conscious will and responsibility: A tribute to Benjamin Libet*. Oxford: Oxford University Press.

Soon, C. S., Brass, M., Heinze, H.-J., & Haynes, J.-D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, *11*, 543–545.

Strawson, P. F. (1974). *Freedom and resentment, and other essays*. London: Egmont Books.

Suzuki, K., Lush, P., Seth, A. K., & Roseboom, W. (2019). Intentional binding without intentional action. *Psychological Science*, *30*(6), 842–853.

Tempesta, D., Couyoumdjian, A., Moroni, F., Marzano, C., De Gennaro, L., & Ferrara, M. (2012). The impact of one night of sleep deprivation on moral judgments. *Social Neuroscience*, *7*(3), 292–300.

Terbeck, S., Kahane, G., McTavish, S., Savulescu, J., Levy, N., Hewstone, M., & Cowen, P. J. (2013). Beta adrenergic blockade reduces utilitarian judgment. *Biological Psychology*, *92*(2), 323–328.

- Tortorolo, P., Monti, J. M., & Vanini, G. (2019). Neurochemistry and pharmacology of sleep. In E. Murillo-Rodríguez (Ed.), *The behavioral, molecular, pharmacological, and clinical basis of the sleep-wake cycle* (pp. 45–83). Cambridge, MA: Academic Press.
- Travers, E., Khalighinejad, N., Schurger, A., & Haggard, P. (2020). Do readiness potentials happen all the time? *NeuroImage*, *206*, 116286.
- Volkow, N. D., Tomasi, D., Wang, G. J., Telang, F., Fowler, J. S., Logan, J., . . . & Ferré, S. (2012). Evidence that sleep deprivation downregulates dopamine D2R in ventral striatum in the human brain. *Journal of Neuroscience*, *32*(19), 6711–6717.
- Vyazovskiy, V. V., Olcese, U., Hanlon, E. C., Nir, Y., Cirelli, C., & Tononi, G. (2011). Local sleep in awake rats. *Nature*, *472*(7344), 443–447.
- Waller, R. R. (2019). *Recent work on agency, freedom, and responsibility: A review*. White Paper. Conshohocken, PA: John Templeton Foundation.
- Waller, R. R. (forthcoming). Science of free will: Neuroscience. In J. Campbell, K. Mickelson, & V. A. White (Eds.), *Blackwell guide to free will*. New York: Wiley-Blackwell.
- Watson, G. (1996). Two faces of responsibility. *Philosophical Topics*, *24*, 227–248.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wegner, D. M., & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, *54*(7), 480–492.
- Whitney, P., Hinson, J. M., Satterfield, B. C., Grant, D. A., Honn, K. A., & Van Dongen, H. P. (2017). Sleep deprivation diminishes attentional control effectiveness and impairs flexible adaptation to changing conditions. *Scientific Reports*, *7*(1), 1–9.
- Yang, L., Lei, Y., Wang, L., Chen, P., Cheng, S., Chen, S., . . . & Yang, Z. (2018). Abnormal functional connectivity density in sleep-deprived subjects. *Brain Imaging and Behavior*, *12*(6), 1650–1657.
- Yoshie, M., & Haggard, P. (2013). Negative emotional outcomes attenuate sense of agency over voluntary actions. *Current Biology*, *23*(20), 2028–2032.
- Yoshie, M., & Haggard, P. (2017). Effects of emotional valence on sense of agency require a predictive model. *Scientific Reports*, *7*(1), 1–8.
- Zhao, R., Zhang, X., Fei, N., Zhu, Y., Sun, J., Liu, P., . . . & Qin, W. (2019). Decreased cortical and subcortical response to inhibition control after sleep deprivation. *Brain Imaging and Behavior*, *13*(3), 638–650.
- Zohar, D., Tzischinsky, O., Epstein, R., & Lavie, P. (2005). The effects of sleep loss on medical residents' emotional reactions to work events: A cognitive-energy model. *Sleep*, *28*(1), 47–54.

This is a section of [doi:10.7551/mitpress/12611.001.0001](https://doi.org/10.7551/mitpress/12611.001.0001)

Neuroscience and Philosophy

Edited by: Felipe De Brigard, Walter Sinnott-Armstrong

Citation:

Neuroscience and Philosophy

Edited by: Felipe De Brigard, Walter Sinnott-Armstrong

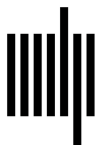
DOI: 10.7551/mitpress/12611.001.0001

ISBN (electronic): 9780262367332

Publisher: The MIT Press

Published: 2022

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

© 2022 The Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif by Westchester Publishing Services. .

Library of Congress Cataloging-in-Publication Data

Names: Brigard, Felipe de, editor. | Sinnott-Armstrong, Walter, 1955– editor.

Title: Neuroscience and philosophy / edited by Felipe De Brigard and
Walter Sinnott-Armstrong.

Description: Cambridge, Massachusetts : The MIT Press, [2022] |

Includes bibliographical references and index.

Identifiers: LCCN 2021000758 | ISBN 9780262045438 (paperback)

Subjects: LCSH: Cognitive neuroscience—Philosophy.

Classification: LCC QP360.5 .N4973 2022 | DDC 612.8/233—dc23

LC record available at <https://lccn.loc.gov/2021000758>