Beyond Button Presses: The Neuroscience of Free and Morally Appraisable Actions

This paper addresses three questions: What are the types of action at issue in the free will and moral responsibility debate? Are the neuroscientists who make claims about free will and moral responsibility studying those types of action? If not, can the existing paradigm in the field be modified to study those types of action? Section one outlines some claims made by neuroscientists about the inefficacy of conscious intentions and the implications of this inefficacy for the existence of free will. Section two argues that, typically, the types of actions at issue in the philosophical literature require proximal or distal conscious decisions (or at least non-actively acquired intentions) and have the right kind of connection to reasons. Section three points out that neuroscientists are not studying this class of actions, as their studies focus on simple commanded actions (e.g., finger or wrist flex) and simple Buridan choices (e.g., push the left or right button). These types of actions do not require conscious control and do not have a connection to the participants' justificatory or motivational reasons for action beyond complying with the experimenter's instructions. Finally, section four then argues that neuroscience already has the resources to study the type of action relevant for free will and moral responsibility and outlines two experiments which focus on skilled actions and moral choices that could be run using the available technology. These kinds of experiments would better address the empirical question about whether conscious intentions and deliberation involving reasons ever play a role in the production of actions that are typically considered to be free and subject to moral evaluation.

1. Neuroscience of Free Will

In 1983, Benjamin Libet and his colleagues used neurological experimental methods to question an arguably key assumption of commonsense

"Beyond Button Presses: The Neuroscience of
Free and Morally Appraisable Actions" by Robyn Repko Waller,
The Monist, vol. 95, no. 3, pp. 441–462.

accounts of agency: do conscious intentions to act cause subsequent (overt) action? Libet and his colleagues found that when participants were asked to perform a series of either preplanned or spontaneous wrist flexes as they felt the intention to do so[1] and report the timing of their awareness of their intention to flex, EEG readings showed brain preparation for the flexing, termed the readiness potential or RP, prior to the participants' awareness of their intention to flex (1983). Libet et al. took these findings to indicate that unconscious brain processes decide (form an intention) to flex, with conscious intentions to flex temporally later and inefficacious. They further generalized these results to all intentional actions, claiming that conscious intentions to A (where A is some overt action) never cause A-ing. Other theorists, such as Daniel Wegner (2002), have argued for similar contentions, using Libet et al. (1983) as support.

Some psychologists and neuroscientists have coupled this claim about the inefficacy of conscious intentions with the further claim that free will exists only if conscious intentions at least sometimes cause our intentional actions. From these claims they argue that neuroscience supports the conclusion that free will does not exist, despite our intuitions to the contrary. For example, Libet himself has argued in some papers that empirical work suggests that there is no free will under the traditional view of free will, only 'free won't' (Libet 2005).[2] Libet reaches this conclusion via the following route: Under a traditional view of free will, which Libet sometimes seems to suggest is an incompatibilist agent-causal or perhaps an incompatibilist event-causal view,[3] a conscious intention to A should appear before or simultaneously with the RP for A-ing; that is, the conscious intention should "command the brain to perform the intended act" (Libet 2005, 553). The fact that participants' reported consciousness of their intention to flex occurred a third of a second after the onset of the RP supports, according to Libet, the proposition that at least some actions that we intuitively take to be free actions are unconsciously initiated. Furthermore, Libet judges that the type of action studied in his experiments, a simple self-initiated finger or wrist flex, is analogous to all intentional bodily actions with regard to RPs preceding conscious intentions to act. While he notes that the timing of conscious intentions to RPs has not been studied in the production of complex actions, he points out that RPs have been evidenced to precede more complex actions. This similarity is enough, according to Libet, to ground general-

izations about the causal production of action for all intentional bodily actions, regardless of complexity (Libet 2005, 560). Thus, given that RPs precede awareness of intention to act, Libet concludes that there is no conscious free will. He does, however, hold open that there may be a conscious 'free won't', an overriding of the already initiated causal process leading from RP to action (Libet 2005). Libet believes that this lack of free will rules out holding individuals morally responsible for the appearance of any intention or urge to perform a morally blameworthy action (and presumably, a morally praiseworthy action—although he does not discuss this). However, Libet also notes that if individuals at least sometimes practice free won't, vetoing the preparation for an action before it's performed, then individuals can be held morally responsible for their overt bodily actions (1999, 54).

Several neuroscientists have replicated Libet's finding that unconscious processes related to later action precede awareness of proximal intentions to act. These authors have also offered their prognosis for the existence of free will in light of these results. For instance, in a paper boldly titled "Unconscious Determinants of Free Decisions in the Human Brain," Soon and his colleagues (2008) asked participants to decide, at a time of their choosing, to either push a button located on the left with their left hand or push a button on the right with their right hand and then report the timing of their awareness of their decision. Using fMRI of the participants' brain activity during the task, Soon and his colleagues demonstrated that brain activity in the prefrontal and parietal cortex up to 10 seconds before the participant pressed a button reliably predicted (above 50% chance) whether the participants would press the left or right button. In contrast, the participants reported that they became aware of their decision to press either the left or right button at about 1 second prior to pushing the button. Soon and his colleagues claim that this substantial gap in time between early related brain activity and decision to act is indicative of "the operation of a network of high-level control areas that begin to prepare an upcoming decision long before it enters awareness" (2008, 543). Thus, they conclude that unconscious activity in the prefrontal and parietal cortex "specifically encodes how a subject is going to decide" (2008, 545), and—given the title—unconsciously determines free decisions. Criticisms have been made about the leap from their data to the conclusion that participants' choice of the left or right button is deter-

mined 10 seconds prior to action (Mele 2012); nevertheless, Soon et al. (2008) is yet another example of a neuroscience study that makes a strong claim about the relationship between free will and conscious intentions on the basis of scientific data.

Two other studies, Lau, Rogers, and Passingham (2007) and Banks and Isham (2009), claim to support Libet's original conclusion that conscious intentions do not initiate intentional action via evidence that the perceived timing of a conscious intention to act in Libet-style studies can be manipulated post-action. Participants in Lau et al. (2007) pushed a button at a time of their choosing and then reported the timing of their awareness of their intention to press the button. In half of the trials, participants were stimulated with transcranial magnetic stimulation (TMS) at the presupplementary sensory motor area (pre-SMA), either immediately after they pressed the button or 200 ms. later. In the other half of the trials sham TMS was applied to participants post-action. The TMS and sham TMS trials were given in random order so that participants would not know whether they were receiving TMS until after pressing the button. Lau and his colleagues found that when participants' pre-SMA was stimulated with TMS at either point after the participants pressed the button, the perceived time of intention to act was shifted backwards in time in comparison to the baseline perceived time of intention to act from the sham TMS trials. Similarly, Banks and Isham (2009) also reported shifts in perception of first awareness of intention to act on the basis of post-action events. Banks and Isham instructed participants to press a button at a time of their choosing and report the timing of their awareness to press the button. The button was held in a box so that participants could press but could not see the button. A computer then randomly generated a beep 5, 20, 40, or 60 ms. after the button had been pressed. This beep was seemingly, to the participants, to signify that the button had been fully pressed. Banks and Isham found that participants' reports of first awareness of intention to press systematically moved forward in time; that is, the longer the beep was delayed after the button press, the closer to 0 (the actual time of the button press) the participants thought the intention to press occurred.

The basic motivation for both studies is the idea, defended by Wegner (2002), that if conscious intentions to A don't actually cause subsequent A-ings, then one explanation of our strong intuition of efficacious conscious intentions might simply be the close temporal association of

(consciously) intending to A and A-ing. As Lau et al. (2007) state, "One strong demonstration for the case of illusory conscious control would be that our perceived temporal order of intentions and actions are, in fact, false. If intentions, in fact, arise after actions, they could not, in principle, be causing actions" (81). Thus, the hypothesis is that although conscious intentions and the actions that they are about are extremely close in time—down to tens or hundreds of milliseconds—the actions are in some sense contributing to the intentions; the intentions don't cause the actions. Armed with their results, Lau and colleagues contend that "the commonsensical view . . . that the main function of experience of intention is for the conscious control of action . . . cannot account for the data presented here" (2007, 89) because of the contribution of post action events to that experience.[4] Likewise, based on their manipulation of perceived time of intention by delayed feedback, Banks and Isham speculate that "the intuitive model [of a conscious intention causing action] has it backwards; generation of responses is largely unconscious, and we infer the moment of decision from the perceived moment of action (Eagleman, 2004)" (2009, 20).

These three recent studies combined with Libet et al. (1983)'s much cited results and conclusions are key examples of neuroscientific claims about the threat of the inefficacy of conscious intentions to folk views of action production and to the existence of free will and moral responsibility. The question is, however, whether the types of actions tested in these studies are the types of actions relevant for free will and moral responsibility. Is Libet (2001) justified in generalizing his results about the role of conscious intentions in producing finger flexes to their role in producing any intentional action? The next section will summarize the types of actions discussed in theoretical work on free will and moral responsibility, and section three will then compare the types of actions in these neuroscientific studies to the ones in the theoretical examples.

2. Free and Morally Appraisable Actions

Philosophers, when discussing free action and actions for which agents can be held morally responsible, typically focus on intentional actions.[5] Although the variety of philosophical accounts of intentional action differs widely (e.g., causal accounts versus noncausal accounts), a common thread to most accounts is that the agent's action must be connected to some set

of the agent's mental states or events in the right way. The state or event at issue, depending on the particular theory of intentional action, is typically one (or more) of the following: desire, belief, intention, decision or choice, or the physical realization of one of these.

Building upon accounts of intentional action to specify the conditions under which an agent freely A-s, philosophers typically discuss one of the following: (1) a case in which agent is uncertain of how to act in a given situation S, deliberates about which course of action to execute, decides to A (i.e., forms an intention to A), and A-s on the basis of that decision; (2) a case in which agent is certain of how to act in a given situation S, nonactively acquires an intention to A,[6] and A-s on the basis of that intention, or (3) if the theorist does not discuss free actions in terms of 'intention', an agent recognizes a reason to A (broadly construed) in S—for example, having a Davidsonian belief-desire pair (Davidson 1963) could satisfy recognizing a reason to A—and A-s for that reason. Note that this list is not meant as a compilation of necessary and sufficient conditions for free action, but rather simply highlights the fact that acquiring, forming, recognizing, and acting on the basis of intentions and/or reasons is a common theme in discussions of free and morally appraisable action.

Among recent theories of free will and moral responsibility, for instance Kane (1996), an incompatibilist about free will (and moral responsibility) and determinism, emphasizes the role of deliberating and deciding in the face of uncertainty about what to do in the production of free action: Kane requires that for an action to be free and one for which an agent is morally responsible, the agent must possess ultimate responsibility for the action such that the action at issue must either be a self-forming action (SFA) or appropriately linked to a past SFA, where SFAs are "regress-stopping, undetermined actions . . . in the life histories of agents" (1996, 124). These SFAs occur during moments of indecision when the agent possesses a conflicting will regarding how to act. Kane notes that SFAs "include such things as choices, decisions, judgments, formation of intention, and efforts of trying . . . SFWs [SFAs] of each kind are motivated by desires and other inclinations; *they involve the formation, alteration, or sustaining of intentions or beliefs that guide action*" (emphasis added) (125). Two types of SFAs that Kane discusses, prudential and moral decisions, are especially relevant to the fourth section of this paper, which outlines a type of neuroscience experiment that investigates the timing of conscious intentions in relation to RP when

participants make a moral choice to perform a moral or prudential action. According to Kane, moral and prudential decisions involve a conflicting agential will in that, prior to both types of decisions, there is a conflict

> between what an agent believes ought to be done and what the agent wants or desires to do. In the moral case, the oughts express moral obligations that are in conflict with self-interested desires. In the prudential case, they concern future or long-term interests that conflict with desires for present or near-term satisfactions. (126)[7]

Kane's main example of a moral decision involves a businesswoman who is hurrying to an important meeting when she notices a person who needs help. The businesswoman must either act in her own self-interest and hurry to her meeting or follow her moral imperative to help the person in distress.

Fischer and Ravizza (1998)'s theory of reasons-responsiveness, a semicompatibilist theory of free will and moral responsibility,[8] also relies heavily on the appropriate connection between reasons and action and on deliberation prior to decision. They hold an account of moral responsibility (and free will) such that "an agent is morally responsible for an action insofar as it issues from his own, moderately reasons-responsive mechanism" (Fischer and Ravizza 1998, 86), in which reasons-responsiveness is explained in terms of the mechanism's regular reasons-receptivity and weak reasons-reactivity. The former requirement refers to "the [mechanism's] capacity to recognize the reasons that exist [to perform an action]," while the latter requirement refers to "the capacity to translate reasons into choices (and then subsequent behavior)" (Fischer and Ravizza 1998, 69). One of Fischer and Ravizza's first examples of a case of free action is one remarkably similar to Kane's moral choice example (although, of course, it does not require the presence of indeterminism):

> We can contrast such cases [where the agent is not acting freely]—in which some responsibility-undermining factor operates—with cases in which there is a *"normal" unimpaired operation of the human deliberative mechanism.* When you deliberate about whether to give 5 percent of your salary to the United Way and consider reasons on both sides, and your decision to give the money is not induced by hypnosis, brainwashing, direct manipulation, psychotic impulses, and so forth, we think you can be legitimately praised for your charitable action (emphasis added). (36)

Note that Fischer and Ravizza, when providing cases of agents acting freely, typically—though not exclusively—cite a normal, conscious deliberative

process involving reasons for acting one way or another as the type of mechanism that issues in free action.[9]

Finally, in addition to both compatibilists and incompatibilists focusing on deliberation and intention formation in the production of free actions, Mele, an agnostic about the compatibility question, also features deliberation and decisions in his account of free will. Mele's agnostic autonomism (Mele 2006) is a disjunctive thesis of two sets of sufficient conditions for autonomy (free will) and hence for moral responsibility, one compatibilist and the other incompatibilist, plus the belief that at least some actual agents are autonomous (free). Both sufficient conditions include the condition that "an agent A-s freely if he nondeviantly A-s on the basis of a rationally formed deliberative judgment that it would be best to A" (200–201) plus other conditions.

In a more recent paper discussing scientific claims about the nonexistence of free will, Mele also highlights the importance of conscious intentions for free will, noting that "If all behavior were produced only by nonconscious processes, and if conscious choices or intentions and their neural correlates were to play no role at all in producing any corresponding actions, free will would be in dire straits" (2010, 169). Baumeister, a social psychologist, agrees with Mele on the importance of effective conscious intentions for the existence of free will, arguing that

> if there are any genuine phenomena associated with the concept of free will, they most likely involve conscious choice. Such a view has to contend with the now widespread belief that consciousness is a useless, feckless epiphenomenon, and that all behavior is guided by nonconscious processes. (2008, 76)

Indeed, it is this thesis—that free will exists only if conscious intentions at least sometimes play a role in the production of action—that neuroscientific work on free will claims to be testing. In the following section, I will take a closer look at types of actions studied in extant Libet-style experiments in light of these examples of theories of free will and moral responsibility to determine whether neuroscientists are looking at the right type of action to test this thesis.

3. Simple Actions and Arbitrary Choices

Recall that participants in Libet et al. (1983) were instructed to perform a series of either preplanned (i.e., at a preset time) or spontaneous (i.e.,

whenever they felt the urge) wrist flexes during the experiment. One salient aspect of this task is that it requires no learned skill to perform; most people can flex their wrists with ease. However, the simplicity of the wrist-flexing task should raise alarms for anyone looking to ascertain the role of conscious proximal intentions from this study.[10] If flexing one's wrists doesn't require any learned skill, participants in Libet et al.'s study could perform these actions automatically—without any conscious attention to doing so. After all, some agents can perform learned skilled actions without consciously intending to do so. Marcel, a neuroscientist, points outs a common example: expert tennis players often intentionally position themselves to make particular shots in response to their opponents' returns without being aware that they intend to so position themselves (2003, 61). Hence, if individuals can learn to perform complicated skilled tasks without consciously doing so, why should we expect Libet et al.'s participants to consciously proximally intend to flex their wrists? It seems that initiation and execution of such a task would be delegated to unconscious brain processes, given the lack of attention required. Perhaps, for example, participants in the study form a conscious distal intention to comply with the instructions, and control is then handed, so to speak, to the unconscious action circuits.[11]

One reason to suppose that participants have conscious proximal intentions to flex in Libet et al. (1983) is that they are instructed to report the timing of their first awareness of their intention to flex and therefore need a conscious item to report.[12] However, even if a conscious proximal intention occurs, it could be the fact that the item is a proximal intention, not that the agent is conscious of it, that contributes to action production. That is, experts performing skilled actions and individuals performing simple actions might require proximal intentions simplicter to execute the action; if such intentions are sometimes conscious, it may not be the participants' consciousness of them that is doing any causal work (Mele 2009, 36). Thus, Libet et al. (1983)'s experimental design is not an adequate test of whether the conscious aspect of conscious proximal intentions plays a causal role in action production.

The types of decisions and actions studied in more recent Libet-style experiments, choosing when to press a button (e.g., Banks and Isham 2009; Lau et al. 2007) and choosing between pressing a left or right button (Soon et al. 2008), are no better than wrist and finger flexes at modeling

the types of decisions and actions that require the presence of conscious intentions. Worse yet, even if we grant for the sake of argument that the experimenter's instructions for participants to report the timing of their first awareness of intention to act produces an effective conscious intention qua its conscious aspect, this resultant conscious intention still lacks any connection to ordinary reasons, moral or otherwise, for acting one way or another and to any need for deliberation prior to action. That is, unlike, say, Kane's businesswoman who is faced with two conflicting reasons for acting one way or another—a selfish desire to be on time to her meeting and a duty to help a person in need—participants in Libet-style experiments have no reason to decide one way or the other: the left button is just as attractive a button to press as the right and pressing the button now is just as attractive an option as pressing the button 50 ms. later.

In *Effective Intentions* (2009), Mele levels this exact criticism against Libet-style experiments that purport to bear upon free action and moral responsibility. Mele notes that such experiments place participants in Buridan scenarios, situations in which "agents are indifferent between or among their leading options" (2009, 83). He points out that unconscious tie-breaking mechanisms may be employed without the agent's awareness in such situations in order to choose a course of action (83). For instance, one Libet-style study discussed in Mele (2009), Brass and Haggard (2007), involves on each trial either proximally intending to press a button that one has already prepared to press or proximally intending not to press a button that one has already prepared to press (vetoing that earlier intention). What reason would a participant have to intend to press the button on any given trial, except for her estimate that she hadn't done so in quite a few trials (perhaps the experimenter instructed her to vary her responses)? This estimate and intention may, we can imagine, be produced without the participants' awareness. Mele goes on to argue that if the agent has a reason to pay attention to her intention to do A or B (for example, she is instructed to report the timing of her awareness of her intention) then that agent consciously decides but nonetheless decides arbitrarily (2009, 84). That is, she arbitrarily chooses to push the button or not push the button. An arbitrary choice (intention formation) is far from the types of choices one makes in the theoretical examples of morally appraisable action—e.g., deciding to benefit oneself instead of helping another, deciding to satisfy a short-term goal instead of satisfying a long-

term goal. This Buridan scenario feature of Libet-style experiments leads Mele to the following conclusion about the experiments upon which many scientists base their claims about free will and moral responsibility for all intentional actions:

> To the extent that Libet studies free will, he studies it in the sphere of proximal decision making in Buridan situations or situations of a similar kind. Generalizing from results obtained in this domain to a view about distal decisions made about important issues in situations of a very different kind would be bold, to say the least. (2009, 85)

Mele, noting Libet's claims that his results are of major import to work on the nature of free will, concludes that

> If this "discovery" has had a profound impact on how some people view free will, that impact rests on an error. That, in certain Buridan-like settings, potential relatively proximal causes of proximal intentions or decisions to A arise unconsciously . . . is a cause neither for worry nor for enthusiasm about free will. (2009, 87)

The Libet-style experiments, then, do not appear to have the evidentiary force that neuroscientists have been claiming: First, the simple movement at issue in these studies is not an adequate test of whether the conscious aspect of conscious proximal intentions plays a causal role in action production. However, that very issue is at the heart of the bold scientific conclusions regarding our lack of free will and limited moral responsibility. Second, the arbitrary free choice afforded participants in the experiments, the choice of when or whether to perform a simple movement, is disconnected from participants' everyday justificatory or motivational reasons—moral, prudential, or otherwise—for action and thus fails to capture the type of decisions and actions for which agents are typically held morally responsible. In order to get better evidence about whether deliberation and conscious proximal intentions (qua conscious) ever play a role in the production of free actions, especially morally appraisable ones, neuroscience will have to tweak the original Libet paradigm.

4. Testing Neuroscientific Claims

As it turns out, testing the efficacy of conscious proximal intentions in producing what are typically held to be free and morally appraisable actions is not a far-off goal of neuroscience but rather is feasible given

existing neuroscience technology and knowledge. In fact, these needed types of experiments can be accomplished with only minor modifications on past experiments. Below is an outline of how these kinds of experiments connect to extant neuroscience studies and what form such experiments might take.[13]

4.1 Conscious proximal intentions and skilled actions

My present concern is primarily with the role of consciousness in the production of intentional action. This discussion, however, will set the stage for some deeper issues about morally appraisable actions (see 4.2). A good place to start when addressing the neglected question of whether an agent's consciousness of her proximal intentions plays a causal role in the production of intentional action is the body of research on skilled action sequences. Recall how agents can intentionally carry out learned skilled action sequences, such as tennis moves, without being aware of any intention to do so. Everyday experience reveals a related phenomenon: when an individual is first learning a skilled action sequence, she must consciously focus on executing the components of that action in order to successfully do so. Sports, with their display of skilled bodily movements, provide an abundance of examples: think of the novice golfer who is learning how to tee off (swing the club) or the gymnast who is learning a back flip. Mastery of all of these athletic actions seems to require repetitive practice involving conscious focus on the body's subtle position and movements. Hence, it seems plausible to say that a novice must form a conscious proximal intention to do the component movements in these action sequences in order to succeed. If this is indeed the case, then skilled action sequences are a good candidate for study in Libet-style experiments aimed at testing the efficacy of conscious proximal intentions in action production.

A survey of the neuroscience literature on attention and skilled action demonstrates that there are skilled action sequences adaptable for study in a Libet-style experiment. In one influential study Passingham (1996) had participants learn a series of finger movements: Participants were instructed to repeat a specified sequence of eight key presses, receiving feedback on their attempts in the form of two tones, corresponding to correct and incorrect. The participants were also asked to do a second task that required conscious attention simultaneously, to generate verbs out

loud. Passingham judged the participants to have learned the sequence—i.e., able to execute the sequence automatically—when they were able to execute both tasks successfully at the same time. This methodology has become the paradigm for research on skilled action sequence learning (e.g., Wu, Kansuku, and Hallett 2004; Jueptner et al. 1997). Learning to successfully execute a key sequence in skilled action sequence studies, unlike performing the tasks in the extant Libet-style studies, *requires* conscious control over one's movements. Hence, there is a plausible counterfactual dependence of the successful completion of the complex finger sequence for a novice and the presence of a conscious intention: If the participant did not consciously proximally intend to perform each step of the finger sequence in the set order (or at least consciously proximally intend to press the keys in a certain order), then she could not have done so successfully. Thus, a Libet-style experiment that compares the timing of a novice's awareness of her proximal intention to execute the key sequence to the timing of the onset of her RP that precedes her executing it would potentially illuminate whether the conscious aspect of conscious proximal intentions to act plays a causal role in action production.

No study as far as I know has tested the timing of conscious proximal intentions in the production of skilled action sequences. However, such a study could easily incorporate elements of a Libet-style study with a Passingham-style novel key sequence task: Suppose that just prior to the study participants are introduced to several key press sequences which are given labels (e.g., sequence 'A' is 'up arrow', 'down arrow', 'left arrow'; sequence 'B' is 'right arrow', 'left arrow', 'down arrow'), and then commanded in a random order to perform any one of the distinct key sequences.[14] While performing a series of these commanded novel key sequences, participants could monitor the timing of the onset of their (single) conscious intention to press the sequence of keys in the set order (e.g., a conscious intention to press 'up', 'down', and then 'left' after hearing the command to execute sequence A)—a conscious proximal intention the participant should plausibly require to successfully execute the task—and report this time after completing the key sequence with a Libet-style clock. In addition, electrodes could be attached to the participants' scalps to measure electrophysiological data of the cerebral activity during the series for each participant, which can be averaged into readiness potentials. Furthermore, a convenient way is available to check whether the presence of

conscious proximal intentions during the series is not due to—or at least not solely due to—instructions to report the timing of the onset of these items: Only half of the participants could be asked to report the timing of the onset of their consciousness of their proximal intention. The other half—those who do not receive any instructions before the series commences regarding awareness of their proximal intentions—should, if asked in an exit interview, confirm that they too experienced consciously intending to press the particular keys of the commanded sequence in the set order when so commanded. Such a check is simple to carry out on the researcher's behalf and would lend credence to the claim that if participants' conscious proximal intentions precede the onset of RPs, the conscious proximal intention is causing the key presses at least partially in virtue of the conscious aspect of the intention.

If in fact the timing of the awareness of proximal intentions to perform a commanded sequence precedes the onset of RPs for participants who had been instructed to report the timing of their first awareness of their intention, these findings would be a step towards vindication of the view of intentional action that Libet et al. rejected in their 1983 study. Such results would support the claims (1) that a conscious intention to act now sometimes appears before any unconscious preparation to act and (2) that conscious intentions (qua conscious) at least sometimes cause subsequent action.[15] These findings, then, would also be a counterexample to Wegner's contention that the consciousness of an intention to act never plays a causal role in the production of action (2002, 55).

4.2 Conscious proximal intentions and moral actions

Although Libet-style experiments that involve novices executing skilled actions are a good bet for ascertaining the timing and role of conscious intentions in the action production stream, someone might object to the use of commanded complex finger sequences on the following grounds: Some researchers might worry that participants who are commanded to perform an action are not intentionally doing so, given that they did not choose to perform the action (see Brass and Haggard [2008] for an argument to this effect). There appears to be two motivations for this worry: (1) there is no uncertainty regarding when or how to act and (2) the action is externally cued and so not internally cued by the agent. However, it seems that one can intentionally act even under instruction as to when and how to do so. For example, (American) football

players plausibly act intentionally when they run a prescribed route at a prescribed time. So, if this intuition is correct, agents acting on command are at least sometimes acting intentionally. Furthermore, Mele has pointed out that philosophers of action would typically consider at least some commanded actions to be intentional actions. The reason for the difference in the use of 'intentional' between researchers like Brass and Haggard and philosophers, Mele argues, is that at least some philosophers of action hold that one can acquire an intention to A and A without deciding (consciously forming an intention) to A (2008, 109). Examples of such intentional actions include habitual actions, such as unlocking one's car door. This type of study may also draw two further worries: First, a commanded task leaves no room for the agent to deliberate about—consider reasons for and against—when, whether, and how to act. Second, a complex finger sequence, although more complex than a button press or wrist flex, is not an action for which we typically hold the agent morally responsible, unless the execution of that finger sequence has a certain type of consequence or is morally blameworthy or praiseworthy under another description. Given that the types of actions at issue in the free will and moral responsibility literature are often preceded by deliberation and are actions according to which we evaluate the agent, the lack of these features in the experiment might seem unsatisfactory.

A second type of experiment, which addresses the above worries, is also within the reach of neuroscience technology. This kind of experiment focuses on a moral choice, one that involves uncertainty about what to do and deliberation, and an action with morally praiseworthy or blameworthy consequences. Both Kane (1996) and Fischer and Ravizza (1998), as mentioned in section 2, discuss cases of moral choices and actions in which the agent considers reasons for helping another against reasons for benefiting herself. So, in accord with these paradigms of free and morally appraisable actions, the aim of such a study would be to get evidence about the role of conscious proximal intentions in *an act of charity*.[16]

A charity act study could be run exactly like Soon et al. (2008) with the following additional features: Before the experiment begins, each participant is told to pick a charity out of a list of options that will benefit from the study. The participant is placed in front of a computer with the following on-screen set-up: a box in the center that represents the amount of money to be given away on that trial, a box on the left that represents the participant's account—how much money she currently has—and a box

of the right that represents the charity's account. The participant then completes a series of trials in which varying amounts of money appear in the center box, and the participant must decide—at no preset time—whether to give that amount of money to herself or to the charity. Noticing when she made the decision to give the money to a particular account, she then must either press the left button to give the money to herself or press the right button to give the money to the charity (i.e., decide on and execute a left or right button press, as in Soon et al. [2008]). Following the button press, the participant indicates when she was first aware of her proximal decision. The next trial will then begin with a new amount in the center box and with the participant's and charity's account boxes reflecting the participant's previous decision. The participant and chosen charity will actually receive the final money total in their respective boxes at the end of the trials; thus the participant's choices are not merely hypothetical ones and therefore better approximate actual moral choices in everyday life. To ensure that participants deliberate before most trials, they could be instructed prior to commencing the study that they must give money to each account at least once and that they are to avoid forming distal policies before the experiment starts of, say, giving the money to each account every other time.[17] Like in Soon et al. (2008), the brain activity of the participants could be monitored with fMRI during the task to examine when certain regions are encoding preparation to push a certain button (left or right). Additional brain regions associated with moral aspects of the task, such as moral reasoning, will be relevant as well.[18] If the participants' first awareness of their proximal intention to give the money to a particular account precedes activity in the brain region responsible for generating a left or a right button press, these data would provide strong evidence against the claim that conscious proximal intentions never play a role in the production of intentional action, particularly paradigmatically free and morally appraisable action. Although deliberation prior to free and morally appraisable actions is often lengthier than the time frames in this type of moral choice study, such a study would still provide valuable evidence about the outcome of the deliberative process, a moral choice.[19]

This act of charity experiment utilizes a Soon et al.-style set-up to gather evidence about moral deliberation and actions and therefore displays several advantages over Libet et al. (1983), Brass and Haggard (2007), Lau et al. (2007), Banks and Isham (2009), and Soon et al. (2008): First,

each trial contains actual uncertainty as to what to do, either give the money to oneself or to charity (or at least the possibility of uncertainty, given that some participants may not follow the instructions to avoid forming distal policies). Second, that uncertainty is not about choosing one of two equally attractive options, but rather about choosing between options of dynamic attractiveness. For instance, if a participant gives the money to himself on the first three trials, he might notice the large discrepancy in the accounts in favor of himself and, as a result, feel guilty. Now, we can imagine, the previously less attractive option of giving to charity may look more attractive in light of the growing motivational strength of his desire to see himself as a humanitarian. Comparatively, his desire, say, to buy a new mobile phone may then lose some motivational strength or suddenly be disregarded. Thus, each trial of the experiment plausibly involves deliberation about which account should get the money on the basis of competing reasons for helping others and for benefiting oneself; that is, the type of action at issue in the experiment has a tight connection to reasons for action and deliberation.

The feasibility of Libet-style experiments on novel skilled actions and Soon et al.-style experiments on moral choices suggests that the types of actions at issue in theories of free will and moral responsibility can be tailored to the types of movements and time frames amenable to neuroscience. Indeed, these types of experiments require only minor alterations of previously conducted studies in neuroscience (specifically, of Passingham [1996]; Soon et al. [2008]). These studies provide an opportunity to establish a connection between moral and skilled actions and the primacy of conscious intentions in action production, and therefore a chance to test neuroscientific claims about the inefficacy of conscious intentions and the relevance of this purported inefficacy for free will and moral responsibility.[20]

Robyn Repko Waller

Florida State University

Notes

1. Libet and his colleagues actually referred to this item in their instructions as the broader disjunction ‘intention’, ‘urge’, or ‘desire’ to flex (1983, 627), but I will refer to this item as an intention for continuity with the rest of the paper.

2. In an earlier paper, Libet appears more optimistic about the existence of free will: "My conclusion that free will, one genuinely free in the non-determined sense, is then that its existence is at least as good, if not a better, scientific opinion than is its denial by determinist theory. Given the speculative nature of both determinist and non-determinist theories, why not adopt the view that we have free will (until some real contradictory evidence may appear, if it ever does). Such a view would at least allow us to proceed in a way that accepts and accommodates our own deep feeling that we do have free will" (1999, 56).

3. The following excerpts from Libet's work are supportive of this interpretation of his take on traditional views of free will: "Are freely voluntary acts subject to macro-deterministic laws or can they appear without such constraints, nondetermined by natural laws and 'truly free'?" (2005, 551); "I myself proposed an experimental design that could test whether conscious will could influence nerve cell activities in the brain, doing so via a putative 'conscious mental field' that could act without any neural connections as the mediators (Libet 1994)" (Libet 1999).

4. Mele (2009) points out that what is really at issue is the timing of the intention to press the button or perhaps the timing of the conscious intention to press, *not* the participants' beliefs about the timing of those events. He notes that the data from Lau et al. (2007) are compatible with it being that case that neither the timing of the intention to press nor the timing of the conscious intention to press is affected by the post button press TMS; only the participants' belief reports about those times may be affected (121). This criticism holds for some of the claims made by Banks and Isham (2009) about the implications of their results as well.

5. A comprehensive theory of free will and moral responsibility covers not only actions but also intentional (and some unintentional) omissions and consequences of those actions and omissions (e.g., Fischer and Ravizza's theory of reasons-responsiveness, [1998]). Here this class will be referred to as simply 'actions' for short.

6. A nonactively acquired intention is one that is not acquired through an act of deciding. This distinction is motivated by the view that decisions are only made when an agent is uncertain about what to do, but that in at least some situations in which an agent lacks uncertainty, the agent still acquires an intention to perform an action (Mele 2009, 8). Mele gives the example of the habitual action of intentionally opening one's office door in the morning. Although one does not decide to unlock the door, one intends to unlock it.

7. C.A. Campbell (1957) also focuses on a conflict between duty and desire as the core feature of free action for which agents are held morally responsible.

8. Semicompatibilism is the view that "moral responsibility is compatible with causal determinism, even if causal determinism is incompatible with freedom to do otherwise" (Fischer and Ravizza 1998, 53).

9. Fischer and Ravizza argue that although their account most naturally lends itself to cases in which an agent acts from a reflective practical reasoning mechanism, their account does not exclude moral responsibility for actions issuing from nonreflective mechanisms, such as habit and instinct (1998, 85). Hence, they contend that acting from a nonreflective mechanism is consistent with recognizing reasons (87).

10. Proximal intentions to A are intentions to A now. Distal intentions to A are intentions to A sometime in the future (Mele 2009, 10). A similar distinction is also made by Bratman (1984) (present-directed intentions and future-directed intentions) and Pacherie (2006) (P-intentions and F-intentions).

11. Roessler and Eilan (2005, 43) make a similar point about Libet et al. (1983).

12. Mele also makes this point (2009, 35).

13. In what follows, I am not proposing hypothetical empirical studies in order to argue for a particular set of conditions according to which it can be said that an agent acts freely and can be held morally responsible for her action. Rather, the revised Libet- and Soon et al.-style studies are meant as examples of how neuroscience can collect better evidence about the processes (e.g., conscious intention formation and moral deliberation) that are typically associated with free will and moral responsibility.

14. The sequences and their labels could have a meaning within the context of a task (e.g., moves in a rudimentary video game, like 'jump') or simply have no specific meaning beyond order of pressing and have variable labels (e.g., 'A'). The reason a researcher might set up the experiment so that that participants are *commanded* in *a random order* to perform one of a number of introduced novel sequences would be to eliminate any unsolicited pre-command preparation to perform a specific key sequence, preparation which might register in the EEG data. To alleviate worries that the participants are performing at an expert level—i.e., that they no longer require conscious proximal intentions to complete the task—the number of trials could be tailored to the number of times participants typically perform these tasks at a novice capacity.

15. Claim 2, is, of course, not only suggested by the expected anecdotal reports of awareness of conscious proximal intentions by the no report group but also by previous studies (e.g., Passingham 1996).

16. Stephen Kearns suggested a study of this kind in conversation. In what follows, I fill in the details and draw a connection to Soon et al. (2008). Of course, moral choice experiments need not be restricted to charity act studies. One could also devise experiments based on other moral scenarios, such as decisions regarding when and how severely to punish a criminal. I am grateful to an anonymous referee for this last point and suggestion. As the referee points out, such cases would be likely to involve a great deal of deliberation prior to decision, as variation in the facts about a given case will affect the appropriateness of punishment and engage relevant consequentialist and retributivist considerations. One feasible model for this type of experiment might be a Soon et al. set-up like the one outlined below for the charity experiment in which participants are given a quick synopsis of a criminal conviction and some facts of the case in the center box and are then asked to press the left button to choose the sentencing option on the left or the right button to choose the sentencing option on the right. To enhance the stakes of the choices for the participants (and thus plausibly increase the likelihood of deliberation), the participants could be told that their choices will be collected as feedback for a committee that is reviewing state sentencing guidelines.

17. Stephen Kearns has suggested (in conversation) the following alternative designs to maximize the number of trials that are preceded by deliberation: (1) Following each button press, on the next trial the choice is between giving the money to oneself or to a *new* charity. (2) On each trial, the participant chooses to give the money to one of two charities, but the two charities given as options change each trial.

18. Based on work by Greene et al. (2001; 2004), the relevant brain areas activated during deliberation and choice for the charity task will likely depend on whether the decision to give the money to either oneself or charity on that trial is driven by utilitarian or deontological considerations. If driven by the former, then one might expect the dorsolateral prefrontal cortex and inferior parietal lobe to show increased activation during deliberation. If driven by the latter, then increased activation of areas associated with affect, such as ventromedial prefrontal cortex, superior temporal sulcus, and amygdala, is

likely. In contrast, recent research by Kahane et al. (2011) suggests patterns of activation during moral choices will instead hinge on whether the judgments made are intuitive or counterintuitive for participants. If this is correct, one might predict either activation of the rostral interior cingulated cortex if the choice on that trial is counterintuitive or activation of the visual and premotor cortex if the choice is intuitive.

19. It should be noted that this charity act study is concerned with conscious proximal intentions, but past studies have looked at the role of conscious distal intentions, which are characteristic of 'extended deliberation' and distal planning (Bratman 1987, 80), in producing intentional action over longer time frames. As Mele (2009) has pointed out, some extant literature does support the idea that conscious distal intentions are effective in action production. Mele cites Gollwitzer's work on implementation intentions: Gollwitzer and his colleagues have found that individuals who form implementation intentions to accomplish a goal task—conscious distal intentions about how (when and where) to do the task—carry out those tasks significantly more often than individuals who do not form implementation intentions to do so (see Gollwitzer 1996; Gollwitzer and Brandstatter 1997; Gollwitzer 1999.). For example, Sheeran and Orbell (2000) (colleagues of Gollwitzer) asked women who had set a goal of attending a cervical cancer screening to specify when and where they would make an appointment for the screening. Women who had formed implementation intentions specifying how they were going to make their appointment were significantly more likely to attend a screening than a control group of women who also had a goal of attending a screening but did not form implementation intentions (96% vs. 69%).

20. A note of thanks to Alfred Mele, Stephen Kearns, and Bruce Waller for their feedback on numerous drafts of this paper.

References

Banks, William P., and Eve A. Isham 2009. "We Infer Rather Than Perceive the Moment We Decided to Act," *Psychological Science* 20: 17–21.

Baumeister, Roy 2008. "Free Will, Consciousness, and Cultural Animals," in Baer, John, James C. Kaufman, and Roy Baumeister, eds. *Are We Free? Psychology and Free Will*, New York: Oxford University Press.

Brass, Marcel, and Patrick Haggard 2007. "To Do or Not to Do: The Neural Signature of Self-Control," *Journal of Neuroscience*, 27: 9141–45.

———. 2008. "The What, When, Whether Model of Intentional Action," *Neuroscientist*, 14: 319–25.

Bratman, Michael E. 1984. "Two Faces of Intention," Reprinted in Mele, Alfred R., ed. *Oxford Readings in Philosophy: The Philosophy of Action*, Oxford: Oxford University Press, 178–203.

———. 1987. *Intention, Plans, and Practical Reason*, Cambridge: Harvard University Press.

Campbell, Charles A. 1957. *On Selfhood and Godhood*, London: Routledge.

Davidson, Donald 1963. "Actions, Reasons, and Causes," *Journal of Philosophy* 60: 685–700

Eagleman, David M. 2004."The When and Where of Intention," *Science* 303: 1144–46.

Fischer, John M., and Mark Ravizza 1998. *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge: Cambridge University Press.

Gollwitzer, Peter M. 1996. "The Volitional Benefits of Planning," in Gollwitzer, Peter M., and John A. Bargh, eds. *The Psychology of Action: Linking Cognition and Motivation to Behavior*, New York: Guilford, 287–312.

———. 1999. "Implementation Intentions: Strong Effects of Simple Plans," *American Psychologist* 54: 493–503.

Gollwitzer, Peter M. and Veronika Brandstätter 1997. "Implementation Intentions and Effective Goal Pursuit," *Journal of Personality and Social Psychology* 73: 186–99.

Greene, Joshua D., R. Brian Sommerville, Leigh E. Nystrom, John M. Darley and Jonathan D. Cohen 2001. "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293: 2105–108.

Greene, Joshua D., Leigh E. Nystrom, Andrew D. Engell, John M. Darley, and Jonathan D. Cohen 2004. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron* 44: 389–400.

Jueptner, Markus, K. M. Stephan, C. D. Frith, D. J. Brooks, R.S.J. Frackowiak, and R. E. Passingham 1997. "Anatomy of Motor Learning. I. Frontal Cortex and Attention to Action," *The Journal of Neurophysiology* 77: 1313–24.

Kahane, Guy, Katja Wiech, Nicholas Shackel, Miguel Farias, Julian Savulescu, and Irene Tracey forthcoming. "The Neural Basis of Intuitive and Counterintuitive Moral Judgment," *Social Cognitive and Affective Neuroscience*.

Kane, Robert. 1996. *The Significance of Free Will*, New York: Oxford University Press.

Lau, Hakwan C., Robert D. Rogers and Richard E. Passingham 2007. "Manipulating the Experienced Onset of Intention after Action Execution," *Journal of Cognitive Neuroscience* 19: 81–90.

Libet, Benjamin. 1994. "A Testable Field Theory of Mind-Brain Interaction," *Journal of Consciousness Studies* 1: 119–26.

———. 1999. "Do We Have Free Will?" *Journal of Consciousness Studies* 6: 47–57.

———. 2005. "Do We Have Free Will?" in Kane, Robert, ed., *The Oxford Handbook of Free Will*, Oxford: Oxford University Press.

Libet, Benjamin, Curtis A. Gleason, Elwood W. Wright, and Dennis K. Pearl 1983. "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential)," *Brain* 106: 623–42.

Marcel, Anthony 2003. "The Sense of Agency: Awareness and Ownership of Action," in Roessler, Johannes and Naomi Eilan, eds., *Agency and Self-Awareness*, Oxford: Oxford University Press, 48–93.

Mele, Alfred R. 2006. *Free Will and Luck*, Oxford: Oxford University Press.

———. 2008. "Recent Work on Free Will and Science," *American Philosophical Quarterly* 45: 107–30.

———. 2009. *Effective Intentions*, Oxford: Oxford University Press.

———. 2010. "Testing Free Will," *Neuroethics* 3: 161–72.

———. 2012. "Another Scientific Threat to Free Will?" *The Monist* 3: 422–40.

Pacherie, Elisabeth. 2006. "Towards a Dynamic Theory of Intentions," in Pockett, Susan, William P., Banks, and Shaun Gallagher, eds. *Does Consciousness Cause Behavior? An Investigation of the Nature of Volition*, Cambridge, MA: MIT Press, 145–67.

Passingham, Richard E. 1996. "Attention to Action," *Philosophical Transactions: Biological Sciences* 351:1473–79.

Roessler, Johannes and Naomi Eilan 2003. "Agency and Self-Awareness," in Roessler, Johannes, and Naomi Eilan, eds. *Agency and Self-Awareness: Issues in Philosophy and Psychology*, Oxford: Oxford University Press, 1–47.

Sheeran, Paschal and Sheina Orbell 2000. "Using Implementation Intentions to Increase Attendance for Cervical Cancer Screening," *Health Psychology* 19: 283–89.

Soon, Chun Siong, Marcel Brass, Hans-Jochen Heinze, and John-Dylan Haynes 2008. "Unconscious Determinants of Free Decisions in the Human Brain," *Nature Neuroscience*, 11: 543–45.

Wegner, Daniel M. 2002. *The Illusion of Conscious Will*, Cambridge, MA: MIT Press.

Wu, Tao, Kenji Kansaku and Mark Hallett 2004. "How Self-Initiated Memorized Movements Become Automatic: A Functional MRI Study," *Journal of Neurophysiology* 91: 1690–98.