

Libet and the Case for Free Will Scepticism

Tim Bayne
University of Oxford and St. Catherine's College
Manor Road
Oxford OX1 3UJ
United Kingdom
tim.bayne@gmail.com

Forthcoming in R. Swinburne (ed.) *Free Will and Modern Science*. Oxford: Oxford University Press. This version of the paper is only a **draft**. For purposes of quotation please consult the published version.

1. Introduction

Free will sceptics claim that we do not possess free will—or at least, that we do not possess nearly as much free will as we think we do. Some free will sceptics hold that the very notion of free will is incoherent, and that no being could possibly possess free will (Strawson this volume). Others allow that the notion of free will is coherent, but hold that features of our cognitive architecture prevent us from possessing free will. My concern in this chapter is with views of the second kind. According to an increasingly influential line of thought, our common-sense commitment to the existence of free will is threatened in unique ways by what we are learning from the sciences of human agency.

We can group such threats into two categories. One kind of threat purports to ‘undercut’ or ‘undermine’ our reasons for belief in free will. To develop a successful objection of this kind one must first identify the basis on which we believe in free will, and then show that this basis is unlikely to yield true beliefs. Although they are not without interest, undercutting objections have not been at the heart of the contemporary case for free will scepticism. Instead, those who invoke the sciences of agency to motivate free will scepticism typically mount rebutting objections to free will. Whereas undercutting arguments attempt to undermine our evidence for free will, rebutting arguments provide what are alleged to be positive reasons against free will.

This chapter examines what is arguably the most influential rebutting objection in the current literature, an objection that appeals to Benjamin Libet’s studies concerning the

neural basis of agency.¹ Although Libet himself stopped short of endorsing free will scepticism on the basis of his results, other theorists have not been so cautious, and his work is often said to show that we lack free will.² I will argue that Libet's findings show no such thing. However, Libet's experiments do raise a number of interesting and important questions for accounts of free will. In particular, Libet's experiments raise challenging questions about the analysis of the concept of free will. In order to determine whether brain science supports free will scepticism we need not only to understand the relevant brain science, we also need to understand just what the common-sense or folk notion of free will commits us to. As we will see, the latter requirement may be as difficult to meet as the former one is.

2. The Libet paradigm

Let us begin with an overview of Libet's experimental paradigm (Libet 1985, Libet et al. 1983). Subjects are told to perform some simply motor action, such as flexing their wrist, at the moment of their choosing within a specified period of time (say, 20 seconds), and that this action should be performed 'spontaneously'. At the same time, they are instructed to monitor their agentic experiences, and to identify the time at which they were first aware of the 'decision', 'urge' or 'intention'—Libet used these terms interchangeably—to act. Subjects do this by watching a clock face with a dial that rotates rapidly (once every 2560 ms). Libet referred to the judgment of the time of their 'decision' ('urge', 'intention') to act as the 'W judgement'. While subjects were both acting and monitoring their urges (intentions, decisions) to act, Libet used an EEG to measure their neural activity. These measurements revealed preparatory brain activity—what Libet called a type II readiness potential (RP)—prior to the action.³ The critical question in

¹ The literature on Libet's experiments is large and expanding. Useful entry points into it are provided by Banks and Pockett (2007), Gomes (1999), Haggard (2008), Levy (2005), Mele (2009) and the chapters in Sinnott-Armstrong and Nadel (2011).

² For examples of free will scepticism that appeal to Libet's work see Banks and Isham (2011), Hallett (2007), Pockett (2004), Roediger et al. (2008), Spence (2009) and Wegner (2002).

³ Libet distinguishes two types of readiness potentials. Actions that are performed spontaneously (as the actions studied in this experiment were said to be) involve a type II RP, whereas pre-planned actions exhibit what Libet calls a type I RP. Type I RPs can be seen up to 1500 ms prior to the action (Libet et al. 1982; Tevena & Miller 2002).

which Libet was interested concerned the temporal relationship between RP and the content of the subjects' W judgments.

The EEG revealed that the RP preceded the subjects' actions by about 550 ms. However, on average subjects reported that they felt that they had decided to move only 200 ms prior to the action (dating that point to the onset of muscle activity initiating the movement). In other words, there appeared to be a gap of about 350 ms between the RP and the point at which subjects claimed to be aware of their decision (urge, intention) to act. In fact, Libet claimed that this gap was around 400 ms in length, for he argued that subjects were aware of their agentic decisions only 150 ms (rather than 200 ms) before they acted. His argument for this claim appealed to the fact that subjects appear to misjudge the point at which a tactile stimulation is applied to the body by about 50 ms. Although I have doubts about whether this correction is justified, not a great deal turns on this issue in the present context and I will accept it here.

As a number of commentators have pointed out, Libet's paradigm is subject to a number of methodological problems (see e.g. the commentaries on Libet 1985). To take just one example of these problems, Libet's paradigm requires subjects to divide their attention between the position of the clock-face and their own agency. The demand to divide one's attention between two perceptual streams in this way is a notorious source of error in temporal order judgements. Despite these difficulties, Libet's basic findings have been replicated by a number of laboratories using studies that are free of these methodological difficulties.⁴ Although there is some variability between studies, the claim that 'Libet-actions'—that is, simple and (relatively) spontaneous motor actions—involve an RP whose onset precedes the time of the subjects' W judgement by about 400 ms or so is largely undisputed. What is in dispute are the implications of these results for questions concerning free will.

Although Libet took his experiments to put pressure on the folk notion of free will he did not think that they established free will scepticism, for he argued that the gap of 150 ms between the agent's conscious decision and the onset of the action allowed for a kind of free will in the form of conscious veto. However, many theorists have seen in Libet's work the death-knell of free will. In their review of his work, Banks and Pocket (2007: 658) describe Libet's experiments as providing "the first direct neurophysiological evidence in support of [the idea that perceived freedom of action is an illusion]."

⁴ For replication of Libet's basic findings see Haggard & Eimer (1999); Keller & Heckhausen (1990); Lau et al. (2004) and Trevena & Miller (2002).

Unfortunately, few sceptics have said exactly how Libet's data is supposed to undermine free will. The central skeptical worry clearly involves the thought that the neural data reveals conscious decisions to be epiphenomenal, but there is more than one way in which this general concern can be corralled into an argument against free will. The following argument seems to me to be closest to capturing the heart of the sceptical appeal to Libet's results, and I will structure my discussion around it:

- (1) The actions studied in the Libet paradigm are not initiated by conscious decisions but are instead initiated by the RP.
- (2) In order to exemplify free will an action must be initiated by a conscious decision.
- (3) So, the actions studied in the Libet paradigm are not freely willed. [From (1) and (2).]
- (4) Actions studied in the Libet paradigm are central exemplars of free will (as intuitively understood), and so if these actions are not freely willed then no (or at least very few) actions are freely willed.
- (5) So no human actions are freely willed. [From (3) and (4).]

I will refer to this as *the sceptical argument*. The sceptical argument is valid, so if it is to be resisted we need to reject one (or more) of its premises. I will examine the premises in reverse order, beginning with (4).

3. The scope of free will

Are the actions that form the focus of the sceptical argument—'Libet-actions'—paradigm examples of our intuitive notion of free will? Libet himself had no doubts about the answer to this question, for he took himself to have studied an "incontrovertible and ideal example of a fully endogeneous and 'freely voluntary' act" (Libet et al. 1983: 640).

However, not everyone shares this view. Adina Roskies, for example, claims that Libet actions are at best 'degenerate' examples of free will, and suggests that we ought to focus on actions that are grounded in our reasons and motivations if we are interested in 'how awareness and action are related insofar as they bear on freedom and responsibility' (Roskies 2011: 19).

To get to the bottom of this issue we need to characterize the kind of actions that are performed in Libet experiments, and to do that we need a taxonomy for actions. Most fundamentally, we can distinguish *automatic* actions from *willed* actions. This distinction can be roughly mapped on to the distinction between endogenous and exogenous actions (Haggard, this volume), and should be thought of as a distinction between two ends of a continuum rather than a distinction between two discrete categories. Automatic actions

flow directly from the agent's standing intentions and pre-potent action routines. Many of our everyday actions—washing the dishes, answering the telephone, reaching for a door handle—are automatic. Our awareness of various features of our environment together with over-learned action schemas conspire to trigger the appropriate intentions with only the minimal participation of conscious deliberation or decision on the part of the agent.

Willed actions, by contrast, require the intervention of executive processes—they require acts of choice and decision. We can distinguish between different forms of willed agency. Consider the experience of finding oneself in a restaurant confronted by a number of equally attractive—or, as the case may be, unattractive—options on the menu. One needs to make a choice, but it does not matter what one orders. Without an act of will one would, like Buridan's ass, starve to death, but little rides on the content of what one wills. We might call this *disinterested agency*. Exercises in disinterested agency must be distinguished from exercises in *deliberative agency*, which involves the interrogation of one's reasons and commitments. Consider Sartre's case of the young man who must choose whether to look after his aged mother or join the resistance. Here, the function of decision-making is not to select from amongst a range of options between which one is relatively indifferent, but to draw on one's reasons in making a good decision. Should one join the resistance or should one follow the dictates of filial duty? Deliberation might fail to deliver a decisive verdict on this matter—and even if it does deliver a verdict, one need not act on it—but the mere fact that one has deliberated entails that one's agency has a different character from what which it would have had had one not deliberated.⁵

Are Libet-actions automatic or willed? Some theorists have suggested that they lie towards the automatic end of the spectrum (Flanagan 1996). The idea, roughly, is that the decision to flex one's wrist *now* (and in such-and-such a way and at such-and-such a time) can be thought of as an automatic component of a complex, willed action. This complex action starts when the experimental procedure begins and one decides to flex one's wrist *at some point* in the next 30 seconds. Having consciously decided to comply with the experimental instruction, the subject offloads the execution of the motor response to automatic processes, with the result that the Libet-action proper is unconsciously initiated.

⁵ Note that although only deliberative agency involves the active interrogation of the agent's reasons, both disinterested and deliberative agency exhibit what Fischer and Ravizza (1998) dub 'reasons-responsiveness.' Thanks to Neil Levy for this point.

Although it is important to recognize that Libet-actions are embedded in a wider agentic context—a context that includes a conscious decision to produce an action of a certain type within a certain temporal window—I am not convinced that Libet-action are best thought of as automatic. Unlike standard examples of automatic actions, subjects in Libet’s experiments are explicitly required to attend to their own agency, and they do report that they decided (had an urge, intended) to produce the action immediately prior to it. Libet-actions may not be the ‘ideal examples’ of fully spontaneous agency that Libet himself takes them to be, but they do seem to be genuine instances of willed agency nonetheless. But although Libet-actions involve an act of will they do not involve deliberation—at least, not immediately prior to the action. In my terms, they are examples of disinterested agency, for the agent has no reason to flex their wrist at one particular time rather than the other, or to flex it in one way rather than another. Indeed, Libet-experiments are explicitly constructed so as to minimize the rational constraints under which the subject acts. We might think of Libet-actions as manifesting the liberty of indifference.

With the foregoing in hand, let us return to the question of whether Libet-action are paradigms of free will (as we intuitively conceive of it). Are disinterested actions our central exemplars of free will, or does that epithet belong to deliberative actions? Philosophers do not agree on the answer to this question, and the systemic research that would be required in order to settle this dispute has not been carried out. That being said, my suspicion is that Roskies is right to identify the central or core cases of free will—at least, the kind of free will that is most intimately related to moral agency—with deliberation and rational reflection.

But even though Libet-actions might not be paradigms of free agency, it seems clear that they *do* fall within the scope of our pre-theoretical notion of free will. (Indeed, I suspect that common sense is inclined to regard even certain *automatic* actions as manifesting some degree of free will; we will return to this point.) As such, the free will sceptic is perfectly within his or her rights to claim that if Libet-actions—and indeed disinterested actions more generally—are not free then an important component of our common-sense conception of free will would be threatened. In sum, although (4) is unacceptable as stated, the sceptical argument is not thereby rendered impotent, for the question of whether Libet-actions manifest free will is itself an important one. Libet-actions might not qualify as ideal examples of free will, and they certainly do not provide us with the only

form of agency that might be of interest to the neuroscience of free will, but they do provide the free will sceptic with a legitimate target.⁶

4. The initiation of free actions

Let us turn now to the second premise of the sceptical argument:

(2) In order to exemplify free will an action must be initiated by a conscious decision.

We can think of (2) as the ‘conceptual’ step of the sceptical argument, for its plausibility turns chiefly on the contours of our everyday (or ‘folk’) notion of free will.

In order to determine whether (2) is plausible, we need to consider in what sense an event might be said to ‘initiate’ an action. I will work with a distinction between a strong sense of initiation and a weak sense of initiation. In the weak sense of the term, an event (ϵ) initiates an action (α) if and only if ϵ is the point of origin of α . The strong sense of initiation entails the weak notion, but adds the requirement that ϵ must be uncaused. This distinction between two notions of ‘initiation’ leads to two readings of (2):

(WEAK): In order to exemplify free will an action must have its point of origin in a conscious decision.

(STRONG): In order to exemplify free will an action must have as its point of origin a conscious decision which is itself uncaused.

Let us consider first what might be said on behalf of (STRONG). This conception of free will characterizes decisions as ‘unmoved movers’—they are the ultimate point of origin of action beyond which the causal chain cannot be traced. Should this constraint on free will be accepted?

Some incompatibilists might argue that it should. Incompatibilists hold that the truth of determinism—the thesis that a description of the current state of the world together with the laws of nature entails a description of all future states of the world—would rule out the possibility of free will. (Compatibilists, by contrast, hold that there is no impossibility between free will and determinism.) Incompatibilists might be inclined to endorse (STRONG) on the grounds that conscious initiation is one way—indeed,

⁶ The free will sceptic might accept the foregoing, but suggest that neuroscientific models of agency derived from Libet’s experiments have the potential to generalize to deliberative actions. Although it would be premature to dismiss this line of thought, the evidence to date suggests that the RP behaves very differently in the context of deliberative agency (Pockett & Purdy 2011).

perhaps the most straightforward way—in which indeterminism could enter into the structure of free agency. Compatibilists, on the other hand, are unlikely to be attracted to (STRONG) for they deny that free will requires the existence of uncaused causes. So, in order to determine whether (STRONG) should be accepted, we may need to first determine whether the notion of free will—that is, the *folk* notion of free will—is to be understood along compatibilist or incompatibilist lines.

We simply do not know the answer to this question. Systematic research into the structure and commitments of the folk or common-sense notion of free will has only just begun, and the results obtained thus far do not paint a clear picture. Some studies suggest that the folk are predominantly compatibilists, others suggest that the folk are predominantly incompatibilists.⁷ One possibility is that ‘the folk’ do not share a single notion of free will, but that some of the folk are compatibilists and others are incompatibilists. Another possibility is that ‘the’ folk notion of free will contains both compatibilist and incompatibilist strands, each of which can be elicited depending on how the subject is probed (Nichols 2006). We do not as yet know enough about the contours of the common-sense notion of free will to decide between these possibilities. In light of this, any version of the sceptical argument that was commitment to (STRONG) would be hostage to the results of future empirical inquiry.

Leaving the commitments of the folk notion of free will to one side, one might argue that (STRONG) is entailed by the conception of decisions that I introduced in §2, in which I said there that the functional role of a decision is to settle what is unsettled. One might argue that this view entails that decisions cannot have fully sufficient causes, for any decision that had a fully sufficient cause would not need to settle anything—indeed, there would be nothing for it to settle.

Although superficially attractive, this line of thought fails to recognize that decisions settle *psychological* uncertainty (Holton 2006). An agent is required to make a decision only when their current psychological states fail to determine what they will do. However, an agent’s behaviour could be psychologically ‘unsettled’ without being neurally ‘settled’. Decisions lack fully sufficient psychological causes, but this does not entail that they (or their neural substrates) lack fully sufficient neural causes, and a decision can initiate an action even if

⁷ For reviews of this literature see Nahmias et al. (2005) and Nichols (2006).

it has a fully sufficient neural cause. So, (STRONG) is not entailed by the conception of decisions that we have employed.⁸

I have suggested that we have no good reason to embrace (STRONG). What about (WEAK)? Should we require that free actions have their point of origin in a conscious decision?

The first comment to make is that the very thought that an action can always be traced back to a *single* point of origin is open to challenge. Rather than thinking of actions as originating with particular discrete events, we might do better to conceive of them as the outcome of multiple events and standing states, no single one of which qualifies as 'the' point of origin of the action. Just as the Nile has more than one tributary, so too many of our actions might result from multiple sources.

Secondly, to the extent that free actions can be traced back to a point of origin, it is by no means obvious that this point of origin must always be a conscious decision. Consider a thoughtless comment that is uttered on the spur of the moment and without forethought. Despite the fact that such an utterance is not consciously initiated, it would be very natural to hold the person who made it responsible for what they had said, and thus to assume that the notion of free will has some kind of grip in such contexts. But, the objection continues, if that is right, then (2) is too demanding, and freely willed actions need not be initiated by conscious decisions.⁹

There are a number of points one might make in response to this objection. For one thing, the advocate of (WEAK) might deny that automatic actions are genuine examples of free will—or at least, that they are genuine examples of the kind of robust free will required for moral agency. Alternatively, they might allow that automatic actions can manifest free will in some sense, but only in virtue of the fact that they are suitably grounded in *prior* exercises of willed agency. (For example, one might take the thoughtless remark to manifest character traits that have been shaped by the agent's previous conscious

⁸ The only kind of determinism that might threaten the possibility of decision-making would be psychological determinism—that is, a determinism involving the agent's standing psychological states. Decisions settled what is unsettled, and if the agent's actions are fully caused by her standing psychological states then they are settled (whether or not the agent herself is aware of this fact). But Libet's experiments provide no support whatsoever for psychological determinism.

⁹ For discussions of this issue see Arpaly (2003), Levy & Bayne (2004), Smith (2005) and Sher (2009).

decisions.) In light of this thought, we might replace (WEAK) with the something akin to the following:

(WEAK*): In order to exemplify free will an action must either have its point of origin in a conscious decision, or it must be appropriately grounded in a prior action that has its point of origin a conscious decision.

Although this emendation is welcome, it doesn't really get to the heart of the matter. As we noted in §3, Libet-actions are not best categorized as automatic. So, although we may well need to modify (WEAK) in order to find a place for free will within the context of automatic agency, the crucial issue that confronts us in attempting to address the skeptical argument from Libet's results concerns the role played by conscious decisions in the context of *willed* actions. The central point is here is this: Libet-actions, unlike automatic actions, are accompanied by the 'phenomenology of conscious initiation': the agent experiences themselves as deciding to act here-and-now, and—arguably—they experience this decision as the point of origin of the action in question.¹⁰ At the centre of the sceptical argument is the thought that if Libet's RP data are correct then this experience of 'origination' must be an illusion, for Libet's data shows that the action is already underway. We cannot, I suggest, attempt to preserve free will in the context of Libet-actions by attempting to ground it in prior acts of conscious initiation.

Where does this leave premise (2)? I have suggested that (2) can be understood in two quite different ways depending on the notion of initiation that is deployed. On a strong reading of the term, (2) requires that free actions must begin with uncaused causes. Although certain incompatibilist conceptions of free will might be committed to this view, it is an open question how central this view is to the common-sense notion of free will. A weak reading of 'initiate', by contrast, requires only that free actions have their point of originate in a conscious decision. Although even this claim is unacceptable as it stands there is something to it, for I have suggested that agents in the Libet paradigm do

¹⁰ See Horgan (2011) for a somewhat different view of the agentic phenomenology that accompanies Libet-actions. Horgan acknowledges that one would experience oneself as beginning to actively undertake an action at some specific moment in time, but he denies that this phenomenology would involve any representation of the mental state causation of one's behaviour. Instead, he suggests, one would experience oneself as undertaking the behaviour, where this phenomenology of authorship should not be thought to include any experience as of mental state causation. Although the issues raised by Horgan's discussion are important I lack the space to do them justice here.

experience their decisions as the point of origin of those simple motor behaviours that they produce in such contexts. The question is whether this experience is at odds with the neural data that Libet obtained. To answer that question we need to consider the first premise of the sceptical argument.

5. Conscious decisions and the readiness potential

The first premise of the sceptical argument is as follows:

- (1) The actions studied in the Libet paradigm are not initiated by conscious decisions but are instead initiated by the RP.

There are a number of ways in which one might attempt to put pressure on this premise. I will begin by considering the possibility that (1) operates with a false contrast between the agent's decision and the RP. The idea here is that the RP and the agent's decision might be the same event—or at least, that the RP might be the neural basis of the agent's decision.¹¹ If this could be established, then we might conclude that the agent's actions are initiated by both the RP *and* their decision.

On the face of things this proposal might appear to be a non-starter. After all, isn't there a gap of around 350-300 ms between the RP and the W-judgements that subjects make? Given this gap, how could it be an open epistemic possibility that the RP is the same event as the agent's decision? In order to answer this question, we must distinguish two ways in which experiences can be associated with temporal properties. On the one hand, experiences can *represent* temporal properties. For example, one might experience a flash of light as occurring before an explosion. This is the temporal structure of the *contents* of experience. At the same time experience themselves also *have* temporal properties. The experience of the flash and the experience of the explosion will have certain locations in objective time. We might call these properties the vehicular properties of experience.

Libet's experiments are concerned with both forms of temporal structure. If the RP data tell us anything about the temporal properties of mental states, then they tell us about their properties considered as vehicles. They tell us when those experiences occur. However, the W data is best understood in terms of the contents of the agent's

¹¹ There are a number of ways in which one might take mental events to be related to brain neural events. One might take the two kinds of events to be merely correlated with each other, one might take mental events to supervene on or be realized by neural events, or one might take mental events to be identical to neural events. I use 'basis' here in a way that is neutral between these various accounts.

experiences. The W judgement represents the agent's opinions as to when their decision to flex their wrist occurred. However, the W judgement itself might occur at some other time. The agent might (or might not) have access to the temporal location of the W judgement, but if so, they don't have introspective access to it simply in virtue of making a W judgement. This is because introspection provides us with access only to the *contents* of experience (see e.g. Tye 2002). So, we need to understand W in terms of the subject's representation of the time of their decision.

The foregoing provides us with a way in which we can see the assumption that is built into (1) might be false, for there is no a priori reason to assume that the temporal location that is represented in the content of an experiences must be identical to the temporal location of the experience itself—that is, of its vehicle. As Daniel Dennett and Marcel Kinsbourne pointed out (1992), just as there is no a priori requirement that the brain employ spatial properties to represent spatial properties, so too there is no a priori requirement that it use temporal properties to represent temporal properties. With this point in hand, Dennett (1991) suggests that the appearance of a gap between the RP and the W judgement results from a failure to appreciate the fact that we are dealing with two kinds of temporal properties. Perhaps, he suggests, the RP is the neural basis of a judgment whose content is expressed in the W judgment: although this mental state itself occurs 550 ms prior to the action, its content represents it as occurring only 200 ms prior to the action.

There are two points to make about this proposal. Firstly, it sits uneasily with a certain conception of the content of decisions. Arguably, the contents of W judgments are token-reflexive. In other words, the referent of 'now' in the decision to flex one's wrist 'now' is simply the location of that decision itself. So, if the RP is the neural basis of that decision, then one has in fact decided to flex one's wrist 550 ms prior to the action, rather than only 200 ms prior to the action as the subject reports. Secondly, even if this proposal shows that there is no incoherence in the thought that the RP could be the neural basis of the agent's decision, we have no evidence that the RP actually *does* form the neural basis of the agent's decision. At present, this proposal represents nothing more than an intriguing account of how the agent's decisions might be related to the RP.¹²

¹² Furthermore, the fact that the correlation between W-judgements and LRP activity is more robust than is the correlation between W-judgement and RP activity (see below, p. xxx) suggests that LRP activity is more likely than RP activity to function as the neural basis of the agent's decision.

David Rosenthal has attempted to undermine (1) from a slightly different angle. Drawing on his higher-order thought account of consciousness, Rosenthal (2002) argues that Libet's results are not at all surprising. According to the higher-order account, a mental state is conscious only in virtue of the fact that it is the intentional target (or object) of another mental state. On this view, decisions and intentions are conscious in virtue of the fact that the agent is conscious *of* them, where this 'consciousness of' involves a higher-order mental state which represents that one has decided or intended to do such-and-such. Given that this higher-order state is subsequent to the decision itself, it is only to be expected—Rosenthal argues—that W judgments lag behind the RP. Although Rosenthal draws heavily on his higher-order theory of consciousness, his proposal is actually independent of any general commitment to the higher-order treatment of consciousness. One could hold that although certain types of mental states are conscious independently of their being monitored by other mental states, as it happens we are conscious of our decisions and intentions only by monitoring them in a certain way.

What should we make of Rosenthal's account? Some theorists will dismiss it on the grounds that intentions and decisions are intrinsically conscious. On this view, the suggestion that decisions or intentions might be unconscious is no more coherent than is the notion that a pain or an itch might be unconscious—decisions and intentions *just are* events in the stream of consciousness.

I don't think that this objection is at all plausible as far as *intentions* are concerned. Consider first distal intentions—intentions to do something in the medium- to long-term. This morning I formed the intention to spend Christmas in Switzerland, but having formed that intention I promptly put the matter out of my mind and was not conscious of it until now. Undeterred, the advocate of the objection might grant that distal intentions need not be conscious, but they might insist that proximal intentions—intentions to do something in the immediate future—must be conscious. But even that claim seems implausible. Consider what it's like to carry out a routine and over-learned action, such as climbing steps or changing gear, whilst lost in thought. In such cases, one's behaviour is guided by intentions (to move this object over here; to move that object over there) of which one may be unaware. There is little reason to insist that intentions must be conscious even when they are active in governing one's behaviour (Mele 2009).

Decisions, on the other hand, might be a different matter. Certainly our intuitive conception of decisions seems to be strongly wedded to the assumption that decisions are acts that one carries out consciously; they are not mental events of which one may (or may not) become conscious. Of course, one might argue that although folk psychology does assume that decisions are intrinsically conscious, this assumption is up for grabs in the

sense that we can make sense of the thought that certain decisions might be unconscious. I'm not so sure that we can. This isn't to say that there aren't legitimate uses of the term 'decision' according to which decisions can be unconscious. Consider, for example, the fact that cognitive neuroscientists describe the visual system as 'deciding' how to categorize a perceptual stimulus. However, one might want to draw a fairly clear distinction between such 'sub-personal' decisions and the kind of personal-level decisions with which we are interested here. Certainly there is some reason to demand that the kind of decisions connected with agent-level responsibility must be conscious. At the very least, a certain amount of suspicion surrounds the notion of an unconscious decision.¹³

Thus far we have considered two ways in which one might attempt to 'identify' the agent's decision with the RP, and I have suggested that there are certain objections to both accounts. However, these objections are not obviously decisive, and it is possible that one (or both) of these accounts can be patched up. Even so, the free will sceptic might argue, neither account would really provide a vindication of free will. 'The reason for this', the sceptic might say, 'is that it is not the agent's conscious decision *as such* which leads them to act but rather the neural activity which underlies that decision. The mental property *per se* is not causally efficacious; instead, the only causally efficacious properties here are neural properties. The decision is a 'free-rider', and hence there is no room for genuine free will.'¹⁴

This line of thought is one manifestation of what is known as the causal exclusion problem. The worry is that the causal efficacy of mental properties is 'excluded' or 'screened off' by the causal efficacy of the neural properties underpinning them. This problem has generated a small mountain of literature, and there is as yet no consensus on how—if at all—it might be solved (see e.g. Bennett 2008; Kallestrup 2006; Kim 1993). However, this is not a literature that we need engage with here, for the issues raised by the exclusion problem are much more general than those which are raised by Libet's data. All that is required in order to provide some intuitive motivation for the exclusion problem is the assumption that the agent's mental events have a physical basis, and most philosophers of mind endorse that assumption independently of any appeal to Libet's

¹³ If we do allow unconscious decisions then we will need non-introspective tools for identifying such states, and it is not entirely clear what such tools might be available.

¹⁴ This kind of worry is suggested by Haggard's comment that 'although consciousness may be a part of brain activity, consciousness cannot cause brain activity, nor can it cause actions' (this volume: xxx).

experiments. Any comprehensive response to free will skepticism must include a solution to the causal exclusion problem, but it is not incumbent on us to provide that solution here.

At the outset of this section I stated that there are two ways in which one might attempt to resist (1). Thus far we have explored the idea that the RP might function as the neural basis of the agent's decision, and hence that both the RP and the agent's decision might be said to initiate the action in virtue of this single event. However, let us assume that the RP is not the neural basis of the agent's decision, and that the RP occurs before the agent decides what to do. Would this fact show that the agent's action is initiated by the RP rather than by his or her decision?

Not necessarily. To see why, we need to return to the question of what it is for an event to initiate an action. I suggested in §4 that an event (ϵ) initiates an action (α) only if ϵ functions as α 's point of origin. The notion of a 'point of origin' can be variously understood, but let us say that ϵ functions as α 's point of origin only if there is a robust correlation between ϵ -type events and α -type events, such that in normal contexts there is a high probability that an ϵ -type event will be followed by an α -type event. (The notion of origination requires more than this, but it is implausible to suppose that it requires less than this.) So, if the RP is the origin of the agent's action, then we ought to expect RP events to be 'immediately' followed by the appropriate action, unless something unusual happens (such as the person being struck by lightning). Is this the case?

As Mele (2009) and Roskies (2011) have observed, we simply do not know. Our ignorance on this point derives from the fact that the RP is measured by a process known as 'back-averaging.' Because the RP on any one trial is obscured by neural noise, what is presented as 'the RP data' is determined by averaging the data collected on a large number of trials. In order to compute this average, the EEG recordings on different trials need to be aligned, and this requires some fixed point—such as the onset of muscle activity or some other observable behaviour on the part of the subject—that can be identified across trials. Any RPs that are not followed by an action simply won't be measured, and so we don't know how robust the correlation between the RP and Libet-actions is.¹⁵

¹⁵ Some commentators have also worried that because the Libet experiments involve averaging across a number of trials, certain aspect of the data might be statistical illusions. In other words, features of the relationship between (say) the RP and the W judgement might characterize the averaged data even though they do not characterize any of the

We do, however, have indirect reasons for thinking that the relation between the RP and subsequent action may not be as tight as that which would need in order to say that the RP is the point of origin of the action. Firstly, we know that the nature of the experimental context can significantly affect both the temporal properties and the strength of the RP signal. Subjects who are highly motivated to perform the task produce a large RP, whereas the RP almost disappears in subjects who have lost interest in the task (McCallum 1988; Deecke et al. 1973; see also Rigoni et al. in press). Secondly, it is possible to make willed responses to stimuli in very much less than 550 ms, which indicates that a type II RP is not ‘the’ point of origin even where it occurs. Thirdly, another neural event—the *lateralized* readiness potential (LRP)—appears to be more strongly coupled to agency than the (generalized) RP is. Whereas the (generalized) RP is symmetrically distributed over both hemispheres, the LRP is restricted to the hemisphere contralateral to the hand that is moved. (In other words, one will see a left hemisphere LRP for a right-handed action and a right hemisphere LRP for left-handed actions.) As Haggard points out (this volume), the evidence indicates that the LRP is more robustly correlated with the subsequent action than the RP is (see Haggard & Eimer 1999). However, the LRP is also very tightly coupled to the W judgments that subjects make, and thus a version of the Libet argument in which (1) is replaced with a corresponding claim about the LRP does not possess even the surface plausibility that (1) does.

Taken together, these three points suggests that the RP is unlikely to qualify as ‘the’ point of origin of the action. If the RP has a psychological interpretation—and it is far from clear that it does—then we should perhaps think of it as the neural realization of quite cognitive and motivational features that contribute to agency. We might think of the RP as the neural basis of an ‘urge’ or ‘inclination’ to act, rather than as the neural basis of the decision to act now (Gomes 1999; Mele 2009). It is one of the many tributaries that contribute to the formation of the action, but it is not ‘the origin’ of the action in any intuitive sense of that term.¹⁶

individual trials that contribute to that grouped data. For discussion of this issue see Roskies (2011) and Trevena and Miller (2002).

¹⁶ One might argue for a similar account of the data reported in Soon et al. (2008), in which the researchers were able to predict which of two decisions agents made up to 10s before they acted by measuring activity in prefrontal and parietal cortex. This neural activity clearly contributes to the agent’s decision, but it is far from clear that it ‘initiates’ the action.

Let me recap the argument of this section. I began by considering responses to (1) which argue that it is possible that Libet-actions could be initiated by both the agent's decision and by the RP, for it might turn out that the RP is the neural basis of the agent's decision. I suggested that although this line of thought should not be dismissed, various considerations weigh against it. I then examined a second, and more plausible, objection to (1)—namely, that the very means of measuring the RP prevents us from determining the robustness of the correlation between it and whether the agent acts. However, various indirect considerations suggest that the correlation between the RP and Libet-actions is not robust enough for us to be justified in describing the RP as the origin of the Libet-action. Instead, the RP may be no more than one of several elements, each of which contributes to the production of the Libet-action.

6. Conclusion

In this chapter I have examined the standard—and arguably most powerful—version of the argument for free will skepticism based on the results of Libet's experiments. I began with the fourth premise of the sceptical argument, and the question of whether Libet-actions qualify as a legitimate target for the scientific investigation of free will. Even though it is doubtful that they are the ideal exemplars of free will that Libet takes them to be, they do fall within the scope of those actions that we intuitively regard as manifesting free will. The discovery that Libet-actions are not freely performed would not itself show that *none* of our actions are freely performed, but it would go some way towards vindicating free will skepticism.

I then turned to the second (or 'conceptual') premise of the sceptical argument, which claims that freely willed actions must be initiated by conscious decisions. I began by distinguishing a strong notion of conscious initiation from a weak notion. The strong notion requires that initiating decisions are uncaused, whereas the weak notion imposes no such requirement, although it does require that initiating decisions have no fully sufficient psychological causes. I suggested that although the precise content of the folk notion of free will is open to debate, it is doubtful whether the folk are committed to the claim that freely willed actions must be consciously initiated in the strong sense of that notion. I also noted that there are questions concerning both the requirement that freely actions must be consciously initiated, and the requirement that free actions must be directly initiated in an act of will.

Finally, in §5 I turned to the first (or 'empirical') premise of the sceptical argument: the claim that Libet-actions are not initiated by conscious decisions but are instead initiated by the RP. We saw that there are two ways in which one might put pressure on this

premise: by arguing that it might be possible to identify the RP with the neural basis of the agent's decision, and by arguing that the RP is merely a contributing factor for the relevant action rather than its point of initiation. We saw that although the first response to (1) is problematic, there is much to be said in favour of the second line of response. All in all, rumours of the 'death' of free will are, I have argued, greatly exaggerated.

What *would* constitute a neurally-based objection to free will? This is a surprisingly difficult question to answer, and a lot will depend on just what the ordinary, intuitive conception of free will it committed to. That said, here is one line of evidence that *might* place our intuitive commitment to free will under strain. Suppose that one found evidence of a neural state that one knew functioned as the neural realization of a psychological state (ϕ) where ϕ occurs immediately prior to the agent's decision. Further, suppose that one had independent reason to think that ϕ was a fully sufficient cause of the agent's decision. Such a discovery, I suggest, would be at odds with the agent's sense of herself as deciding what to do—as 'making up her mind.' What one would have discovered is that the agent was not in the state of psychological uncertainty that she took herself to be in. She experienced herself as 'making up her mind', but in fact her mind was already 'made up'. For what it's worth, my hunch is that the sciences of human agency are exceedingly unlikely to provide us with evidence along these lines, but that's a matter for another occasion.

Although my central focus in this chapter has been with the sceptical challenge to free will, I do not want to give the impression that this challenge constitutes the only—or even the most important—point of contact between the neuroscience of agency and questions of free will. Rather than asking whether the sciences of human agency undermine our commitment to free will, we might instead look to them for insights into the nature of free will. What is it about human cognitive architecture that provides us with the capacity for free agency? How can the domain of human freedom be expanded? How can impediments to the exercise of free will be removed? With few exceptions (see e.g. Holton 2009, Roskies 2010), philosophical engagement with the sciences of agency has been dominated by attempts to address the sceptical challenge. That focus has been understandable, but perhaps it is time for us to consider how neuroscience might enrich our understanding of free and autonomous agency.¹⁷

¹⁷ I am very grateful to Neil Levy, Richard Swinburne and an anonymous reviewer for their helpful comments on earlier versions of this paper.

References

- Arpaly, N. 2003. *Unprincipled Virtue*. New York: Oxford University Press.
- Banks, W. P. & Pockett, S. 2007. Benjamin Libet's work on the neuroscience of free will. In M. Velmans & S. Schneider (eds.), *The Blackwell Companion to Consciousness*. Blackwell.
- Banks, W.P. and Isham, E. A. 2011. Do we really know what we are doing? Implications of reported time of decision for theories of volition. In W. Sinnott-Armstrong and L. Nadel (eds.) *Conscious Will and Responsibility*. New York: Oxford University Press, pp. 47-60.
- Deecke, L., Becker, W. Grözinger, B., Scheid, P. & Kornhuber, H.H. 1973. Human brain potentials preceding voluntary limb movements. In W.C. McCallum & J.R. Knott (eds.) *Electroencephalography and Clinical Neurophysiological Supplement: Event-related Slow Potentials of the Brain: Their Relations to Behavior (Vol. 33)*, Elsevier: Amsterdam, pp. 87-94.
- Dennett, D. 1991. *Consciousness Explained*. Brown and Little.
- Dennett, D. & Kinsbourne, M. 1992. Time and the observer. *Behavioral and Brain Sciences*, 15: 183–247.
- Fischer, J. M. and Ravizza, M. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.
- Flanagan, O. 1996. Neuroscience, agency, and the meaning of life. In *Self-Expressions*. Oxford: Oxford University Press.
- Gomes, G. 1999. Volition and the readiness potential. *Journal of Consciousness Studies*, 6/8-9: 59-76.
- Haggard, P. 2006. Conscious intention and the sense of agency. In N. Sebanz and W. Prinz (eds.) *Disorders of Volition*. Cambridge, MA: MIT Press, 69-86.
- Haggard, P. 2008. Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience*, 9: 934-46.
- Haggard, P. and Eimer, M. 1999. On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research* 126, 128–133.
- Haggard, P., Newman, C. & Magno, E. 1999. On the perceived time of voluntary actions. *British Journal of Psychology*, 90: 291-303.
- Hallett, M. 2007. Volitional control of movement: the physiology of free will, *Clinical Neurophysiology* 118: 1179-92.

- Holton, R. 2006. *The act of choice*. *Philosophers' Imprint*, 6 (3): 1-15
- Holton, R. 2009. *Willing, Wanting, Waiting*. Oxford: Oxford University Press.
- Horgan, T. 2011. The phenomenology of agency and the Libet results. In W. Sinnott-Armstrong and L. Nadel (eds.) *Conscious Will and Responsibility*. New York: Oxford University Press, pp. 159-72.
- Horgan, T., Tienson, J., & Graham, G. 2003. The phenomenology of first-person agency, In S. Walter and H-D Heckmann (eds) *Physicalism and Mental Causation: The Metaphysics of Mind and Action*. Exeter, UK: Imprint Academic (pp. 323-40).
- Kallestrup, J. 2006. The causal *exclusion argument*. *Philosophical Studies*, 131 (2): 459-85.
- Keller, I. & Heckhausen, H. 1990. Readiness Potentials preceding spontaneous motor acts: voluntary vs. involuntary control. *Electroencephalography and Clinical Neurophysiology*, 76: 351-61.
- Kim, J. 1993. The non-reductivist's troubles with mental causation. In J. Heil and A. Mele (eds.) *Mental Causation*. Oxford: Clarendon Press.
- Lau, H. C., Rogers, R.D., Haggard, P., Passingham, R.E. 2004. Attention to intention, *Science*, 303/20: 1208-1210.
- Levy, N. 2005. Libet's impossible demand. *Journal of Consciousness Studies*, 12: 67-76.
- Levy, N. & Bayne, T. 2004. Doing without deliberation: automatism, automaticity, and moral accountability, *International Review of Psychiatry*, 16/4: 209-15.
- Libet, B. 1985. Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8: 529-566.
- Libet, B., Gleason, C.A., Wright, E.W. & Pearl, D. 1983. Time of unconscious intention to act in relation to onset of cerebral activity (Readiness-Potential), *Brain*, 106: 623-42.
- McCallum, W.C. 1988. Potentials related to expectancy, preparation and motor activity. In T. W. Picton (ed.) *EEG Handbook Vol. 3: Human Event-Related Potentials*. Amsterdam, Elsevier, 427-534.
- Mele, A. 2009. *Effective Intentions: The Power of Conscious Will*. New York: Oxford University Press.
- Miller, J. O., Vieweg, P., Kruize, N., & McLea, B. 2010. Subjective reports of stimulus, response, and decision times in speeded tasks: How accurate are decision time reports? *Consciousness & Cognition*, 19: 1013-36.

- Nahmias, E., Morris, S.G., Nadelhoffer, T., & Turner, J. 2005. Surveying freedom: Folk intuitions about free will and responsibility, *Philosophical Psychology*, 18/5: 561-84.
- Nichols, S. 2006. Folk intuitions on free will. *Journal of Cognition and Culture*, 6/1-2:57-85.
- Pockett, S. 2004. Does consciousness cause behaviour? *Journal of Consciousness Studies*, 11/2: 23-40.
- Pockett, S. and Purdy, S. 2011. Are voluntary movements initiated preconsciously? The relationships between readiness potentials, urges, and decisions. In W. Sinnott-Armstrong and L. Nadel (eds.) *Conscious Will and Responsibility*. New York: Oxford University Press, pp. 34-46.
- Rigoni, D., Brass, M. & Sartori, G. In press. Inducing disbelief in free will alters brain correlates of preconscious motor preparation, *Psychological Science*.
- Roediger, H. K., Goode, M.K., and Zaromb, F.M. 2008. Free will and the control of action. In J. Baer, J.C. Kaufman and R.F. Baumeister (eds.) *Are We Free?* Oxford: Oxford University Press, pp. 205-225.
- Rosenthal, D. 2002. The timing of conscious states. *Consciousness and Cognition*, 11: 215-20.
- Roskies, A. 2010. How does neuroscience affect our conception of volition? *Annual Review of Neuroscience*, 33: 109-30.
- Roskies, A. 2011. Why Libet's studies don't pose a threat to free will. In W. Sinnott-Armstrong and L. Nadel (eds.) *Conscious Will and Responsibility*. New York: Oxford University Press, pp. 11-22.
- Sher, G. 2009. *Who Knew? Responsibility Without Awareness*. Oxford: Oxford University Press.
- Sinnott-Armstrong, W. & Nadel, L. 2011. *Conscious Will and Responsibility*. New York: Oxford University Press.
- Smith, A. 2005. Responsibility for attitudes: activity and passivity in mental life. *Ethics*, 115: 236-271.
- Soon, C. S., Brass, M., Heinze, H-J., Haynes, J-D. 2008. Unconscious determinants of free decisions in the human brain, *Nature Neuroscience*, 11/5: 543-5.
- Spence, S. 2009. *The Actor's Brain: Exploring the Cognitive Neuroscience of Free Will*. New York: Oxford University Press.
- Trevena, J.A. & Miller, J. 2002. Cortical movement preparation before and after a conscious decision to move, *Consciousness and Cognition*, 11: 162-90.

Tye, M. 2002. Representationalism and the *transparency* of experience. *Noûs*, 36: 137–151.

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