

The Neuroscientific Study of Free Will: A Diagnosis of the Controversy

Markus Schlosser, m.e.schlosser@rug.nl

Forthcoming in *Synthese*

This is the author's copy, which may differ in minor details from the print version

Abstract: Benjamin Libet's work paved the way for the neuroscientific study of free will. Other scientists have praised this research as groundbreaking. In philosophy, the reception has been more negative, often even dismissive. First, I will propose a diagnosis of this striking discrepancy. I will suggest that the experiments seem irrelevant, from the perspective of philosophy, due to the way in which they operationalize free will. In particular, I will argue that this operational definition does not capture free will properly and that it is based on a false dichotomy between internal and external causes. However, I will also suggest that this problem could be overcome, as there are no obvious obstacles to an operationalization of free will that is in accord with the philosophical conception of free will.

Keywords: Neuroscience of free will; definition of free will; Libet experiment; acting for reasons; reason-responsiveness

1. Introduction

Conscious will and free will were for a long time considered to be beyond the reach of rigorous scientific investigation. This has changed quite dramatically in the past few decades. Consciousness and its role in the initiation and guidance of voluntary action is now the subject of much empirical research, and even free will is studied scientifically. This is in large part due to the pioneering work of Benjamin Libet (Libet et al. 1983, Libet 1985, and 1999, for instance). This research paved the way for the scientific study of conscious will and free

will, and it is widely acknowledged as a groundbreaking contribution.¹ But there has also been a lot of criticism. Both scientists and philosophers have criticized parts of the experimental method—in particular, Libet’s method of measuring the onset time of conscious will. Philosophers have also been very critical of Libet’s interpretation of the data and of his conclusions about free will. Their criticism has been considerably stronger than that of other scientists. They have argued that Libet’s conclusions about conscious will and free will do not follow, even if it is granted that the experimental methods and results are sound.² However, the reception of Libet’s challenge to free will has been mixed within philosophy itself. Some philosophers see a need to engage with Libet’s challenge in detail and with care, which suggests that they take the findings seriously.³ Others, however, are more dismissive. They think that Libet’s findings are simply irrelevant to the problem of free will.⁴

This striking discrepancy in the reception of the neuroscientific study of free will is not restricted to Libet’s work. Many philosophers who dismiss Libet’s research as irrelevant to free will are equally dismissive of follow-up studies that develop the Libet paradigm (such as Haggard & Eimer 1999 and Soon et al. 2008). My aim here is to provide a diagnosis of this discrepancy and to offer some suggestions on how to remedy the situation. First, I will suggest that the neuroscientific studies appear to be irrelevant, from the perspective of philosophy, because they deploy an operational definition of free will that fails to capture free

¹ See Libet 2002, where he provides a long list of fellow scientists who have “not only accepted our findings and interpretations, but have even enthusiastically praised these achievements and their experimental ingenuity” (292).

² One can get a good sense of the striking differences in the assessment of Libet’s findings (and related empirical evidence) from a discussion of the topic “Is free will an illusion?” in the *Chronicle of Higher Education* (2012, March 18), which includes contributions from scientists and philosophers.

³ See, for instance, the contributions by philosophers to the peer commentaries to Libet 1985 and to Pockett et al. 2006, Baer et al. 2008, and Sinnott-Armstrong & Nadel 2011. The most detailed engagement with Libet’s challenge by a philosopher is Mele 2009.

⁴ Although I have come across this view many times in conversation with other philosophers, it is difficult to find clear statements in print. The main reason for this is probably that the philosophers who publish academic work on this topic are the ones who take the empirical challenge to free will seriously, whereas the ones who think it is irrelevant do not bother to write about it. However, statements of the view that the evidence is simply irrelevant to free will (or “no threat at all”) are provided by three philosophers on the *BBC Radio 4* program “In our time: philosophy” (2011, March 10).

will properly. In particular, I will argue that this operationalization of free will is based on a false dichotomy between internal and external causes. However, we will also see that the underlying mismatch between the neuroscientific and the philosophical conception of free will is a contingent happenstance. We will see, in other words, that there is nothing—in principle and in practice—that would prevent an operationalization of free will that is in accord with the philosophical conception.

2. The Libet Paradigm

First of all, I will provide a summary account of Libet’s seminal experiment and of two important follow-up studies that developed and improved the paradigm. Libet conducted several experiments and studies on the timing of movement and conscious awareness. In the main experiment—generally referred to as ‘the Libet experiment’—, participants were instructed to sit down and to perform a simple and predefined movement when they felt like doing so. More specifically, they were instructed to let the urge or wish to move “come on its own, spontaneously” (Libet 1985: 531). During this, EEG measurements were taken from the midline of the subject’s scalp, roughly over motor and premotor areas, and EMG signals were taken from the muscles that are involved in the movement. At the same time, participants were looking at a clock-like device and they were instructed to remember and report the position of the clock’s revolving dot at the time when they experienced the “first awareness of wanting to move” (ibid.: 532). Libet called this conscious event ‘W’ and he referred to it variously as a conscious ‘urge’, ‘wish’, ‘choice’, or ‘intention’. This report allowed the experimenters to determine the timing of the onset of W. It was known at the time that voluntary movements are preceded by a brain potential that arises from motor and premotor areas, called the ‘readiness potential’ (‘RP’, for short). In the Libet experiment, the onset of the RP, obtained by EEG over the scalp, occurred on average 550ms before the onset of the movement. The main result of the experiment was that the onset of the RP also preceded W, the reported time of conscious will, on average by 350ms.

The Libet experiment investigated only the very specific choice *when* to perform a predefined movement (‘when-decisions’). However, when we are interested in free will, we are also interested in the question of whether we are free to decide *what* to do (‘what-

decisions’).⁵ Haggard and Eimer (1999) conducted a version of the Libet experiment that addresses this issue. They gave participants a free choice between two actions: pressing a button with either their right or left index finger. Further, they investigated the role of the lateralized readiness potential (‘LRP’). The LRP occurs later than the RP and it is more specific to the ensuing movement. In particular, when the agent performs a movement with a right limb, the LRP is generated by activity in the left hemisphere, and *vice versa*. Haggard and Eimer compared the timing of the onset of W to both the onset of the RP and the LRP. They replicated the main finding of Libet that the RP precedes W, and they found that even the onset of the LRP precedes W (at least, on average).

As mentioned, the accuracy of Libet’s method for measuring the onset times of W has been questioned and criticized many times (see Glynn 1990, Gomes 2002, Joordens et al. 2002, and Danquah et al. 2007, for instance). This issue has been addressed in a more recent experiment conducted by John-Dylan Haynes and colleagues (Soon et al. 2008). In their experiment, the clock-like device was replaced by a screen on which participants see a stream of letters that updates every 500ms. This measurement method is obviously more coarse-grained, as the unit of measurement is almost as long as the whole RP process in Libet’s experiment (550ms). Despite this, Haynes and colleagues were able to replicate the basic result of the Libet experiment. As in the experiment by Haggard and Eimer, participants were given a free choice between two actions (pressing one of two buttons with either their left or right index finger). Using statistical pattern recognition techniques, Haynes and colleagues decoded brain activations from fMRI data that are predictive of both what-decisions and when-decisions. Brain activity that is predictive of what-decisions was found as early as 10s before the conscious choices, and brain activity that is predictive of when-decisions was found up to 5s before the conscious choices. This seems to replicate the main finding in a rather dramatic fashion: conscious choices are preceded by specific unconscious processes not just up to a few hundredth of a second, but up to several seconds.⁶

⁵ Arguably, what-decisions are more significant, because when-decisions are usually subordinate to what-decisions in the sense that we usually decide when to do something in order to decide how to implement a what-decision.

⁶ Two recent studies have raised serious questions concerning the measurement and interpretation of the RP. The experiments in Miller et al. 2011 suggest that effects of the clock-monitoring could be partly responsible for the

3. Conclusions Concerning Conscious Will and Free Will

What did the scientists themselves conclude from the findings? Libet drew two main conclusions; one concerning conscious will and one concerning free will. First, we do not consciously initiate our actions, because the evidence shows that the “initiation of a spontaneous voluntary act begins unconsciously” (1985: 529).⁷ Second, we do not have free will in the sense we commonly think do. We commonly think, according to Libet, that free will requires conscious initiation, and the evidence shows that we do not consciously initiate our actions (especially Libet 1999).⁸

Haggard and Eimer’s main aim was to investigate whether the LRP is more likely to be the cause of W than the RP (1999: 128). They found that the onset times of RPs do not covary with the reported onset times of W, whereas the onset times of LRPs do. They concluded, in conjunction with a principle concerning the temporal co-variation of causes and effects, that the LRP is more likely to be the cause of W. They did not draw any further conclusions about free will. Given, however, that the onset times of LRP also precede W, the Haggard and Eimer experiment may be taken to support the same main conclusions as Libet’s

main finding in the Libet experiment. The experiments conducted by Schurger et al. 2012 suggest that the precise timing of the decision is partly determined by ongoing spontaneous fluctuations of neural activity and that the decision to act might occur much later in the RP process than Libet assumed. Further, experiments conducted by Schlegel et al. (*forthcoming*) raise serious doubts about Haggard & Eimer’s findings concerning the role of the LRP. Given this, it seems no longer credible to hold that the main findings of the Libet experiment are an established empirical fact (as it has often been claimed). However, further experiments are required to establish the truth on those issues, and the mentioned studies do not challenge the results of Soon et al. 2008.

⁷ According to Libet, a conscious intention to initiate a movement “should precede or at least coincide with the onset of the specific cerebral processes that mediate the act” (1985: 529). One has to assume here that by “mediate the act” Libet meant “initiate the act”. Otherwise, the conclusion that the initiation of a spontaneous voluntary act begins unconsciously does not follow. Further, one must assume that the RP is such a “specific cerebral processes” that initiates the act. This assumption, I should mention, has attracted a lot of criticism. See Keller & Heckhausen 1990, Zhu 2003, Mele 2009, for instance.

⁸ Another prominent empirical challenge concerning the role of consciousness in the initiation of action is due to Wegner (2002), who argued that the experience of conscious will is an illusion. On some occasions, Wegner appears to use the terms ‘conscious will’ and ‘free will’ interchangeably, but he does not draw any explicit conclusions about free will. I have addressed Wegner’s challenge elsewhere (Schlosser 2012a).

(see Haggard & Libet 2001).⁹ Furthermore, given that participants in the Haggard and Eimer experiment are given a choice between two options, it seems that their result raises the Libet challenge for when-decisions as well as what-decisions.

The situation is similar with respect to the Soon et al. experiment. Soon et al. did not draw any explicit conclusions about free will. However, in the introduction to their article they mention that the Libet experiment has been taken to suggest that “freedom is no more than an illusion”, because it suggests that “our actions are initiated by unconscious mental processes before we become aware of our intention to act” (2008: 543). Given that their main aim was also to improve the Libet experiment, and given that their experiment replicates the basic result of the Libet experiment, it seems that Soon et al. took themselves to have provided further evidence that corroborates Libet’s challenge to free will.¹⁰

Libet thought also that the evidence from his experiment is compatible with what he called “conscious veto control” (1999: 51–55). During debriefing, some participants reported that they had sometimes chosen not to act on the urge to move—they had aborted or vetoed the urge to act. To investigate this further, Libet conducted a version of the experiment in which participants were instructed to inhibit the movement when the conscious wish arises. During this, EEG signals were recorded that resembled RPs. This suggests, Libet argued, that the time between W and the onset of the movement leaves the opportunity to consciously veto the movement. At one point, Libet seemed to suggest that we should revise our conception of free will accordingly, so that we can say that we have free will *in this sense* (Libet 1999). Later, he and Haggard proposed a different line, suggesting that the evidence is compatible with our having “free won’t” but not with our having “conscious free will” (Haggard & Libet 2001). It is clear, I think, that these suggestions were supposed to weaken the blow, as it were. Despite this, it remains clear that Libet thought throughout that the evidence shows that we do not have free will *in the sense we ordinarily think we do*.

⁹ See, however, Schlegel et al. (*forthcoming*).

¹⁰ However, the findings are controversial due to the fact that the unconscious brain activations predict the conscious choices only with a relatively low accuracy of about 60%. In other words, the decoded unconscious brain activations raise the probability of the subsequent conscious choices to just above the level of chance.

4. Are the Findings Even Relevant to Free Will?

The conclusion concerning free will has been very controversial. The debates about this presume that the evidence is relevant to the problem of free will. However, as mentioned, many philosophers think that even this assumption is too generous.¹¹ In this section, I will propose a diagnosis of this dismissive stance. This, I should point out, will not be an empirical claim about what reasons or thought processes actually underlie the dismissive stance of those philosophers (although I think that the diagnosis will capture this to some extent). Rather, this diagnosis will provide a *rational reconstruction* of the dismissive stance, and it will be both an explanation and a justification. I will suggest, in other words, that there are good reasons to think that the evidence *is* irrelevant to the problem of free will.

Libet's conclusions about conscious will and free will are closely related and both are based on an operational definition which is, as Libet claimed, "in accord with common views" (1999: 47). This definition consists of two conditions. According to the first, free will requires that the choice of the action is endogenous in the sense that "there should be no external control or cue to affect the occurrence or emergence of the voluntary act" (ibid.). Elsewhere, Libet expressed this by saying that a free choice must be "self-initiated" and "not in direct response to an external stimulus or cue" (1985: 531 and 529). According to the second condition, free will requires that the choice and execution of the action should be accompanied by the sense or feeling that one is acting "on one's own initiative" and that one "could control what is being done" (1999: 47).

A first thing to note here is that this definition is rather idiosyncratic, from the perspective of philosophy, in the sense that it defines free will in terms of a sense or feeling of willing. In the philosophical debate, the relation between free will and the conscious experience of will has not been a central issue, and all the philosophical standard positions on free will are silent about the connection between them. However, I do not think that this is the real problem here, and I think that most philosophers would grant Libet the assumption that genuinely free choices must be accompanied by an experience of conscious will. What is problematic, however, is the first condition that the choice of the action must be endogenous or self-initiated in the sense that "there should be no external control or cue to affect the

¹¹ See note 4.

occurrence or emergence of the voluntary act”. All the experiments in the Libet paradigm clearly reflect and implement this condition on free will: they all study choices that are entirely endogenous in this sense. This is why, I think, that the findings from the experiments in the Libet paradigm seem simply irrelevant to many philosophers. In order to explain this, I will first have to say more about the philosophical conception of free will.

In philosophy, virtually every claim about free will has been contested, and virtually every question concerning free will has been controversial. This includes claims and questions concerning the concept, nature, and definition of free will. It is, however, widely agreed within the philosophical debate that questions concerning free will are, first and foremost, questions concerning choices and actions that are *based on reasons*. On most views, this does not mean that the relevant choices and actions must be based on conscious deliberation about pros and cons, and it does not mean that they must be based on normatively good reasons. Rather, to say that the relevant choices and actions are based on reasons means, more modestly, that they can be *rationalized* from the agent’s point of view—that there is, in Davidson’s words, “some feature, consequence, or aspect of the action that the agent wanted, desired, prized, held dear, thought dutiful, beneficial, obligatory, or agreeable” (1963: 685). This constraint on free will is widely accepted, both by compatibilists and by incompatibilists about free will (for a small but representative sampling of references see Dennett 1984, Wolf 1990, and Vihvelin 2004 for compatibilist views, and van Inwagen 1989, Kane 1996, and Clarke 2003 for incompatibilist views).

Why should we think about free will in this way? Giving a full and satisfactory answer to this question would lead us too far astray. But let me identify, in broad outline, three reasons. First, it seems very plausible to assume that only *intentional* actions can be properly free, and it is widely agreed, in philosophy, that intentional action is best explained in terms of acting for reasons (the relevant *loci classici* are Anscombe 1957 and Davidson 1963). Second, it is common in philosophy to contrast free will with the “liberty of indifference” (see Hobart 1934, van Inwagen 1989: 417 and Kane 1996: 109, for instance). The liberty of indifference is the freedom to choose among options that are indistinguishable, such as the freedom to pick one item from a stack of qualitatively identical products. As there is nothing that distinguishes the options, there is no reason to prefer one option over the others, and so the freedom to choose one of them is the freedom to choose with indifference. The point here

is not that this is not real freedom. Rather, the point is that this kind of freedom is insignificant and uninteresting. This, in turn, is related to the third point, which is that only choices that are based on reasons are choices that matter to us. Such choices matter to us precisely because seeing a reason for (or against) a certain alternative is to see something that counts for (or against) it, whereas choices between indistinguishable alternatives do not matter to us precisely because we cannot see any reason to prefer any one alternative over the others.

As just mentioned, in philosophy it is common to contrast free will with the freedom of indifference. Given, however, that the freedom of indifference is a kind of freedom and given that it is a freedom of choice, why should we not say that the freedom of indifference is a kind of free will after all? The dispute here is verbal, at best, and verbal disputes can be solved by means of terminological stipulations. One may take the contrast literally and reserve the label ‘free will’ for choices that are based on reasons. But one may just as well concede that the freedom of indifference is a kind of free will—albeit an insignificant and uninteresting kind of free will. I shall opt here for this second way of resolving the issue, and I shall use the term ‘free will proper’ in order to refer to choices that are both free and based on reasons. Given all this, we can say, then, that it is widely accepted within philosophy that free will proper is the freedom to choose and act on the basis of reasons.¹²

Against this background, it is easy to see why the findings from the experiments in the Libet paradigm have appeared to be irrelevant to philosophers. The experiments, it seems, investigate only the freedom of indifference. This is only an insignificant and uninteresting kind of freedom, and the findings tell us nothing about free will proper. This point holds for all the mentioned studies. The Libet experiment investigates choices about when to execute a predefined movement. There is no reason to perform the movement at a particular point in time rather than another, and there is no reason to assume that participants have preferences about when to perform it. The choice is made with indifference and it seems, therefore, insignificant and uninteresting. The mentioned follow-up experiments investigate decisions about what to do. But they do not get to the bottom of the problem. In both the Haggard and

¹² Note that this provides a necessary condition: the choice of an action is made with free will proper only if it is based on reasons. It is not meant to provide a sufficient condition; it is not meant to give a full definition; and it is not meant to give an explanation of what freedom is.

Eimer and the Soon et al. experiment, participants are given a choice between two actions. It is quite clear, however, that the freedom of choice remains the freedom of indifference. There is no reason whatsoever to press a button with the left index finger rather than the right, and there is no reason to think that participants have any preference about the matter.

It seems, then, that the experiments in the Libet paradigm investigate only the freedom of indifference. This provides a good explanation of why many philosophers dismiss the experimental findings as irrelevant. As indicated, I think also that this diagnosis provides a justification. The claim that the findings are irrelevant seems justified, because an investigation of choices that are made with the freedom of indifference does seem to be irrelevant to free will proper. Given our terminology, this justification is only partial, because it does not establish that the findings are entirely irrelevant to free will. We granted that the freedom of indifference is a kind of free will, and so one may hold that the findings do tell us something about one variety of free will. Given, however, that this is an insignificant and uninteresting kind of freedom, this is only a minor qualification. We can still conclude that the findings are irrelevant to questions concerning free will *proper*.

It has often been observed that the experiments in the Libet paradigm require participants to make highly unusual and artificial choices, which are devoid of any normative significance and personal value, and which appear to be very much unlike the choices that we make in our everyday lives. Given this, it seems that there is no obvious reason to think that the results are generalizable. Further, one might think that this response is not only closely related to the proposed diagnosis, but that it makes essentially the same point. I think, however, that it is important to make a distinction here. This non-generalizability response is based on the observation that there is a striking difference in the *kind* of the choices that are being made, whereas the proposed diagnosis concerns what the choices are *based on*. In particular, the former holds that the results are not generalizable, whereas the proposed diagnosis is not committed to this claim and it does not depend on it. In fact, according to the proposed diagnosis, it may well be that *all* our choices have unconscious precursors. If a choice is not based on reasons, then it may seem plausible to assume that the choice is made with free will only if there are no unconscious precursors that cause the choice. However, if a choice is based on reasons, then one would expect that those reasons causally influence the choice. And this causal connection between the agent's reasons and the choice may well

involve unconscious processes. In other words, even if the neuroscientific findings were generalizable, they would not raise a direct challenge to free will proper, because they tell us nothing about the role of the agent's reasons, and because the influence of the agent's reasons may well be mediated by unconscious processes (for more on this see Schlosser 2012b).

5. Libet's Definition of Free Will

An obvious question that poses itself at this point is this: Why did Libet define free will in a way that is so obviously out of step with the predominant view in philosophy? In social psychology, there is a traditional conceptual framework that distinguishes between two kinds of causal attribution: attributions to the person and attributions to the situation. This is, essentially, a distinction between two different kinds of behavior explanation: explanations in terms of internal causes and explanations in terms of external causes (for an overview and references see Malle 2004 and 2011). Of course, Libet was not a social psychologist. But the attribution framework is widely known and it has been widely deployed. Given this, it may well be that Libet's operational definition of free will was shaped by the attribution framework—and if it was not shaped by it, then it is, at least, perfectly in line with it. The definition says that free choices must be self-initiated in the sense that there is no external factor that affects the choice of the act. It requires, in other words, that free choices have exclusively internal causes, and free choices are thereby contrasted, implicitly, with choices that are influenced by external factors. So, this way of thinking about free will coheres perfectly with the attribution framework, and I think that it is important to note this for the following reason.

Recently, Bertram Malle has criticized the attribution framework for distorting the way in which we think about intentional agency. In particular, Malle has argued that the dichotomy between person and situation, or internal and external causes, “fundamentally misrepresents the properties of reason explanation” (2011: 338). To illustrate this point, Malle compared the following two explanations: “Jack bought the house *because it was secluded*” and “Jill bought the house *because she wanted privacy*” (ibid.). According to the attribution framework, these are two fundamentally different explanations. The first explains in terms of the situation (external cause), and the second explains in terms of the person (internal cause). This is obviously misleading, according to Malle, because the difference between the two

explanations is only superficial—it is merely a difference on the “linguistic surface” (ibid.). To elaborate, it seems clear that someone who gives the first explanation means to imply, or presumes, that Jack thinks or believes that the house is secluded. In other words, someone who gives an explanation in terms of the situation means to imply, or presumes, that there is a corresponding explanation in terms of the person. Given this, it is misleading to divide and classify explanations of behavior in terms of situation vs. person (or external vs. internal), because explanations in terms of the situation presuppose that there is a corresponding explanation in terms of the person (or the person’s mental states and events). This suggests that the distinction at the heart of the attribution framework is a false dichotomy.

This critique of the attribution framework connects neatly with the offered critique of Libet’s definition of free will. Libet’s definition is deficient, as I have argued, because it does not capture choices and actions that are based on reasons. In conjunction with Malle’s critique of the attribution framework, we can now see that any definition of free will that is based on this framework will be deficient in this respect, because the attribution framework cannot capture choices and actions that are based on reasons. This framework distinguishes between explanations in terms of internal and external factors. However, choosing and acting on the basis of reasons involves, typically, *both* internal and external factors. When an agent can freely choose what to do, then it is, in some sense, up to the agent what to do. We may say that a free choice can, in this sense, be explained by internal factors. But if one chooses and acts for reasons, one typically responds to situational or external factors that one takes as giving one reasons for the action. Such external factors are often situational factors, but they need not be. Consider two versions of a simple example. Suppose that you decided to take an umbrella when you left the house. In the first case, you took the umbrella because you noticed dark clouds over the sky. In the second, you took the umbrella because you remembered that the weather forecast had predicted rain. You responded to a situational factor only in the first case. But an external factor played a role in both cases, and when we explain the two actions we assume that you took that factor as giving you a reason for the action. It seems clear, then, that choosing and acting on the basis of reasons is neither a purely internal nor a purely external phenomenon. It has both internal and external components, which is why it seems appropriate to say that we *respond to* reasons when we choose and act on the basis of reasons.

It does seem clear now that the distinction at the heart of the attribution framework really is a false dichotomy and that any definition of free will within this framework will be inadequate.

One might object here that this diagnosis holds only if we assume a certain *externalist* conception, according to which reasons are external facts (Parfit 1997 and Dancy 2000, for instance). According to a widely accepted *internalist* conception, acting for reasons consists in acting on the basis of the relevant desires, beliefs, and intentions (Davidson 1963, Bratman 1987, for instance). On such views, acting for reasons might appear to be a purely internal phenomenon after all. But this is mistaken. First of all, we should distinguish between accounts of *reasons* and accounts of what it is to *act for* reasons. Further, we should distinguish between *strong* externalist views of reasons, according to which external facts can provide an agent with reasons independently of the agent's mental states, and *weak* externalist views, according to which external facts provide an agent with reasons depending on whether the agent has the relevant desires, beliefs, and intentions. The proposed diagnosis is by no means committed to a strong externalist view of reasons, and weak externalist views are fully compatible with internalist accounts of what it is to act for reasons. In fact, even strong externalist views must be compatible with some internalist account of what it is to act for reasons. For even if reasons are facts, in the strong sense, they must in some way affect the agent's mental states in order to motivate intentional action. To illustrate, return to the example: "Jack bought the house because it was secluded". The view that Jack's reason was (the external fact) that the house was secluded must be compatible with some view on how we can explain the action in terms of Jack's mental states (such as the belief that the house was secluded and the desire or intention to live in seclusion). Second, explanations in terms of the agent's mental states are usually not purely internalist insofar as mental states are usually about external states. When we explain Jack's action in terms of his belief, for instance, we explain it partly in terms of what this belief is about: that the house is secluded. Third, standard internalist conceptions are not purely internalist in the further sense that they do not deny that a rational agent must to some extent be responsive to situational or external facts. Internalist conceptions allow for the possibility that choices and actions can be based on reasons even when they are based on false beliefs or seemingly irrational desires. However,

they all deny that an agent's mental states are reason-giving if they are altogether cut off from or insensitive to the relevant external facts (a *locus classicus* is Davidson 1970).¹³

6. The Lesson and Possible Directions for Future Research

To recapitulate, the distinction between internal and external causes is a false dichotomy that distorts our understanding of intentional agency, because it distorts our understanding of what it is to choose and act for reasons. Given this, any operational definition of free will that deploys this distinction is bound to be deficient. It might capture the freedom of indifference, but it cannot capture the freedom that we really care about—freedom to choose and act on the basis of reasons. When applied to free will, the distinction is in fact doubly misleading. First, it draws a boundary where there should be none. Choosing and acting for reasons has both internal and external components, which is, as we have seen, reflected in the properties of reason explanation. Second, when applied to free will, the distinction suggests the view that causation by external factors is incompatible with free will. This explains why Libet defined free will in the way he did, and it would explain why other scientists, such as Patrick Haggard, think that “the concept of free will implies that the participants themselves, rather than the experimental situation, brings the triggers for action” (Haggard 2011: 10). The concept of free will implies this only on a narrow construal of a ‘trigger’. Choosing and acting with free will is incompatible with being triggered to react in a way that bypasses one’s rational capacities. The problem with the passage, however, is that not every external factor (or ‘experimental situation’) triggers a response in this way. As I have point out, an external factor may be something that provides a reason for action. If one responds to such a factor

¹³ See also Fischer & Ravizza’s (1998) influential account of reason-responsiveness. They argue that “reason-receptivity” is relative to subjective features of the agent, but that it cannot be reduced to “the agent’s point of view”. Rather, reason-receptivity must be “grounded in reality” (73). Further, they claim that this view fits with “the widest possible selection of plausible views about reasons for action” (68, note 11). Note that this view is compatible even with source libertarianism, according to which the agent must be the *source* or *origin* of his or her free choices (Kane 1996, for instance). All source libertarians acknowledge that the agent should be influenced by reasons. Given that genuine reason-responsiveness requires at least some sensitivity to external factors, even source libertarians can and should allow the influence of external factors.

because one recognizes that the factor provides a reason, then the response is not a mere reaction. More importantly, such a response may very well be free.

Another way of putting this is to say that being influenced by external factors does not necessarily mean that we are coerced or manipulated (in the commonsensical meaning of ‘being manipulated’). For if the external factor gives us a good reason, we are rationally persuaded. Return to one of our examples. You took the umbrella because you noticed dark clouds over the sky. The action was not a mere reaction, nor were you being coerced or manipulated. For all we know, you acted freely in response to recognizing a reason for action.¹⁴

Given this, we can also see now that there is something wrong with the way in which Libet operationalized the idea that free actions must be consciously initiated. It is a mistake to think that the initiation of actions must be entirely endogenous. We often initiate actions on the basis of reasons. Such actions may well be free, but they are not entirely endogenous, given that they are based on reasons.

It is clear what lesson should be drawn from this: scientific investigations of intentional action and free will should abandon the dichotomy between internal and external causes (or person and situation attribution). This seems unproblematic insofar as there is no apparent theoretical or practical necessity in using the attribution framework. There is, in other words, no obvious reason to think that an operational definition of intentional action or free will must deploy the attribution framework. Using this framework, Libet excluded all external causes

¹⁴ One might think that such an action was free only if the reason did not determine the action and only if the agent could have done otherwise. According to compatibilists, free choices may be causally determined. What matters, rather, is that the choice of the action is not coerced or manipulated. This condition is satisfied: the fact that one is *persuaded* by a factor is incompatible with being coerced or manipulated by that factor. In contrast, incompatibilists would insist that the reason must not determine the choice. This is also compatible with what I have claimed: we may assume that a factor that provides a reason exerts only a causal *influence* on the choice of the action. Is it required that the agent could have done otherwise? This issue has been very controversial (see, for instance, Dennett 1984, van Inwagen 1989, and Kane 1996). Fischer & Ravizza (1998) have argued that reason-responsiveness does not require the ability to do otherwise. Elsewhere, I have proposed that the ability to do otherwise can be analyzed in terms of reason-responsiveness (Schlosser 2012b). Be that as it may. The important point is that nothing hangs on this issue here, for everything that I have claimed here is compatible with *both* the affirmation and the denial of the view that the agent must be able to do otherwise.

and influences. This resulted in an inadequate definition of free will. Given the proposed diagnosis, it seems clear how the operational definition can be improved. First, external factors and influences should be allowed. Second, we should distinguish between *types* of external causes. In particular, external factors that merely trigger or manipulate choices and actions should be distinguished from external factors that influence choices and actions by way of providing the agent with reasons to act. As explained, the causal influence of an external factor does not rule out free will if that factor provides an agent with a reason. To the contrary, if the agent chooses freely what to do in response to something that provides a reason, the choice is made with free will proper.

We know that these points can be implemented, because there exists already a large amount of empirical research that allows and controls the influence of external reason-giving factors. Take, for instance, the behavioral and neuroscientific research on economic decision-making and gambling tasks. In these experiments, participants are presented with various offers and constraints that give them reasons for and against certain options (for an overview to neuroscientific studies see Glimcher et al. 2009, for instance). Or consider the meta-analysis of evidence concerning the efficacy of intentions by Webb and Sheeran (2006). This analysis covers forty-nine studies in which participants are given good reasons for significant real-life choices, such as reasons to take physical exercise, to wear a seatbelt, engage in parent-child communication, to have regular health checkups, to practice safe sex, to quit smoking, and so on. They found that the evidence supports the hypothesis that interventions on intentions by way of giving good reasons engenders the corresponding changes in intentions and actions. The fact that changes in intentions were manipulated by controlled interventions does not rule out free will, because the agents were ‘manipulated’ here only in the scientific sense of the term. They were not manipulated in a sense that is incompatible with free will, because the manipulation consisted in persuasion by reasons.

Of course, all these experiments were not designed for the study of free will, and it is not clear whether they can tell us anything interesting about free will. However, the point here is merely that these experiments study the kind of choices and actions that should be the focus in the study of free will—choices and actions that are based on reasons. The question of what the empirical study of such choices and actions can tell us about free will is largely an open empirical question. The experiments in the Libet paradigm have investigated unconscious

precursors of conscious choices. This has raised several questions concerning the relationship between unconscious brain activity and conscious choices. Does the unconscious brain activity cause the conscious choices? Does it merely prepare or bias the choices? Are the conscious events in question really choices or intentions? Or are they rather urges or desires? And so on. The influence of reasons would raise various additional questions. How is the influence of conscious reasons related to unconscious brain activity? Does the conscious recognition of reasons also have unconscious precursors? If so, does this undermine or threaten free will? And so forth.

However, there are also larger questions in the background that are not empirical questions—at least not in any straightforward sense of the term. Most obviously, there is the traditional philosophical question of whether or not free will is compatible with determinism. How one thinks about this issue may have a significant impact on how one interprets empirical findings about conscious decision-making (more on this below). Further, how one thinks about the relationship between the mind and the brain may affect how one interprets neuroscientific findings, and it may affect how particular scientists design their experiments. Perhaps the mind-body problem can one day be solved by empirical research. However, should an empirical solution to this problem be possible at all, it cannot come about in a conceptual and metaphysical vacuum, as it were. It would presuppose a certain conceptual framework with certain metaphysical commitments.

7. Epiphenomenalism, Determinism, and Mechanism

I have argued that the experimental findings are irrelevant to questions concerning free will proper, simply because they do not investigate free will proper. However, one might think that, even if this is correct, the findings do challenge free will proper, because they raise challenges that generalize to all decisions and actions (including the ones that are based on reasons). One might think, in particular, that they raise a perfectly general challenge because they establish one of the following doctrines: epiphenomenalism, determinism, or mechanism. It is important to address these worries. My replies will have to be short, however, as a detailed discussion is beyond the scope of this paper. Despite this, it should become clear that these challenges are misguided or, at least, far from convincing.

According to the first challenge, the findings raise a general challenge because they establish epiphenomenalism. As pointed out, it seems very plausible to hold that free will proper concerns choices and actions that are based on reasons. Common sense would suggest that consciousness should play a role in this: conscious reasons and conscious intentions should be causally efficacious in the initiation and guidance of such actions. But the experiments in the Libet paradigm suggest that conscious intentions are epiphenomenal (that is, causally inert). They are, therefore, clearly relevant to an issue concerning free will proper.

First of all, it has been argued, many times, that the experiments do not show or even suggest that conscious intentions are epiphenomenal (see Zhu 2003, Mele 2009, and many of the contributions to Pockett et al. 2006, Baer et al. 2008, and Sinnott-Armstrong & Nadel 2011, for instance). Let me mention here only one main point. None of the experiments establishes causation, as none of them features controlled interventions. Libet and Soon et al. establish a temporal ordering. But, of course, the fact that unconscious activity precedes the conscious intention, which is then followed by a movement, does not show that the conscious intention is not a cause of the movement. The Haggard and Eimer experiment establishes also only a temporal co-variation, which suggests only that the LRP is more likely to be the cause of the conscious intention than the RP. Again, this does not even suggest that the conscious intention is causally inert.

Moreover, the claim of irrelevance can be defended against this objection even if we grant that the evidence establishes epiphenomenalism. For if the findings established epiphenomenalism, they would do so only for *proximate* intentions (that is, roughly, the conscious events that proximately precede the actions).¹⁵ They would not establish, or even suggest, that the mental states that are associated with the agent's reasons are epiphenomenal: they would not show that the agent's desires and beliefs are epiphenomenal, and they would not show that the agent's distal intentions are epiphenomenal. In fact, it seems that the

¹⁵ I assume here that the conscious events that proximately precede the movements are conscious *intentions*. I should note, however, that this assumption has also been questioned. For instance, Keller and Heckhausen (1990: 359) suggested that the conscious events in question were the "result of selective attention" to look for an urge to move, which was, in turn, induced by the artificial setup and the instructions of the experiment. They suggested, in other words, that the conscious events in question are neither intentions nor conscious events that precede ordinary actions.

experiments in the Libet paradigm presuppose, like all experiments on human decision-making and voluntary action, that conscious reasons and intentions can be causally efficacious. Participants agree, consciously, to take part in the experiment and to follow the instructions to the best of their abilities. The fact that they usually do what they are instructed to do suggests that the provision of instructions results in the conscious formation of mental states (desires, beliefs, and distal intentions) that initiate and guide their actions during the experiment. Not only does it seem clear that these states are causally efficacious, but the experimental method presupposes that the provision of instructions is efficacious in this way. In other words, the experiments not only fail to show epiphenomenalism for the mental states that are relevant to free will, but they presuppose their causal efficacy (see also Zhu 2003).

Moreover, there is plenty of direct evidence for the efficacy of conscious reasons and distal intentions. For instance, the evidence from the research on implementation intentions shows that the conscious formation of distal intentions is effective in the self-regulation of behavior (Gollwitzer & Sheeran 2006). Similarly, the mentioned meta-analysis of studies on the efficacy of intentions shows that providing participants with good reasons for certain actions tends to bring about the corresponding changes in intentions and actions (Webb & Sheeran 2006).

Libet addressed a related objection when he argued that the empirical study of free will may focus on proximal intentions, because all other mental states “must somehow work their way into a ‘final common motor activation pathway’ in the brain” (1985: 530). The idea here seems to be that the efficacy of proximal intentions is crucial, as they are the final gates for the initiation of voluntary movement, so to speak. This, however, is beside the point in at least two respects. First, it does not address the problem that proximal intentions are formed only with the freedom of indifference if they are not based on reasons. Second, Libet’s reply does not show that the causal efficacy of other mental states depends on the causal efficacy of proximal intentions. The fact remains that the provision of instructions is obviously efficacious, which suggests that the mental states that are consciously formed in response to instructions do, somehow, work their way into the motor control system. We can conclude, then, that the challenge from epiphenomenalism is unsuccessful. It is questionable that the findings establish epiphenomenalism. And even if we assumed that the findings show that

proximate intentions are epiphenomenal, they would not show, or even suggest, that conscious desires, beliefs, and distal intentions are epiphenomenal as well.

According to a second challenge, the experiments generalize to free will proper, because they show that our choices and actions are causally determined. Here we should distinguish sharply between the following two questions. First, is free will incompatible with causal determinism? Second, does the evidence show that our choices and action are causally determined? Libet and Haynes (who conducted Soon et al. 2008) have made it clear that they are incompatibilists: they believe that we do not have free will if our choices and actions are causally determined (see Libet 1999: 55 and Haynes 2011: note 1). It is very controversial, however, whether or not free will really is incompatible with determinism. More importantly, the compatibility question is not an empirical question, and it is clear that Libet's and Haynes' view on this issue is not based on any experimental findings. So, if the evidence is relevant at all, it is not relevant to the compatibility question. Does the evidence show, then, that our choices and action are causally determined? I fail to see how. As already pointed out, the experiments do not even provide clear evidence in support of causation, let alone causal *determination*.¹⁶ This is particularly clear for the Soon et al. experiment, which suggests only that the brain activity raises the probability of a certain choice just above the level of chance.¹⁷ But it holds also for the Libet and the Haggard and Eimer experiments, as they establish only temporal orderings and temporal co-variation. And, again, even if the findings suggested causal determination, the challenge would still depend on the controversial metaphysical assumption that free will is incompatible with causal determinism.

According to a related third and final challenge, the experiments generalize to free will proper, because they show that the causation of our choices and actions is mechanistic: they show that our choices and actions are caused entirely by events and processes at the neuronal level, leaving no room for the conscious self to play an active role in the initiation and guidance of actions. Again, we have to distinguish two issues. Is the mechanistic view incompatible with free will? Does the evidence show that the mechanistic view is true? Concerning the first question, we need to be careful. If the role of the conscious *self* can be understood in terms of the role of conscious mental *states and events*, and if conscious mental

¹⁶ I assume here that causation may be non-deterministic, such that causes raise the probability of effects.

¹⁷ See note 10.

states and events are *realized* by brain states, then it is far from obvious that the mechanistic causation of behavior is incompatible with free will. On a charitable interpretation, the mechanistic view does not exclude the role of the conscious self at all. Rather, it provides a reductive explanation. Very roughly, on this view the role of the conscious self can be understood in terms of the role of conscious mental states and events, and the role of conscious mental states and events is realized by activity in the brain. Of course, some reject this reductionism, but the majority of philosophers and scientists see no problem with a mechanistic account of the mind and agency. In any case, it should be noted that a rejection of the mechanistic view is not itself based on evidence. It is based, rather, on strong metaphysical assumptions concerning the nature and the role of the conscious self. Let us turn, then, to the second question: Does the evidence show that the mechanistic view is true? Again, it is difficult to see how. No one denies that brain activity is necessary for the exercise of conscious will. Given this, it seems clear that the discovery of neural activity that is correlated with the conscious initiation and guidance of voluntary actions is perfectly compatible with a *rejection* of the mechanistic view. For such neural activity may only be necessary, not sufficient, for the exercise of free will in the initiation and guidance of action. Moreover, even if such neural activity was sufficient, this would be a problem only if it could be shown that mechanistic accounts of mind and agency are mistaken—and all the familiar arguments against mechanistic views are, as far as I can tell, either far from decisive or begging the question.

8. Conclusion

I have argued that the findings from the experiments in the Libet paradigm appear to be irrelevant, from the perspective of philosophy, due to a mismatch between the philosophical conception and the neuroscientific operationalization of free will. In particular, this operationalization fails to capture what I called free will *proper*—the freedom to make choices on the basis of reasons. However, I have also argued that this mismatch could be overcome, as there are no obvious obstacles to an operationalization that is in line with the philosophical conception. Given, furthermore, that the philosophical conception of free will does indeed capture the kind of freedom that matters to us, it is clear that this is not just a possibility, but a *desideratum*: the operational definition of free will should be revised in

accord with the philosophical conception, because this conception captures the kind of freedom that we care about.

References

- Anscombe, E. (1957). *Intention*. Oxford: Basil Blackwell.
- Baer, J., Kaufman, J.C., & Baumeister, R.F. (eds.) (2008). *Are We free? Psychology and Free Will*. Oxford: Oxford University Press.
- BBC Radio 4. (2011, March 10). In our time: Philosophy. http://downloads.bbc.co.uk/podcasts/radio4/iotp/iotp_20110310-1023b.mp3. Accessed 15 Oct 2012.
- Bratman, M.E. (1987). *Intentions, plans, and practical reason*. Cambridge: Harvard University Press.
- Chronicle of Higher Education. (2012, March 18). Is free will an illusion? <<http://chronicle.com/article/Is-Free-Will-an-Illusion-/131159/>>.
- Chronicle of Higher Education. (2012, March 18). Is free will an illusion? <http://chronicle.com/article/Is-Free-Will-an-Illusion-/131159/>. Accessed 15 Oct 2012.
- Clarke, R. (2003). *Libertarian accounts of free will*. Oxford: Oxford University Press.
- Dancy, J. (2000). *Practical reality*. Oxford: Oxford University Press.
- Danquah, A.N., Farrell, M.J., O'Boyle, D.J. (2007). Biases in the subjective timing of perceptual events: Libet et al. (1983) revisited. *Consciousness and Cognition*, 17, 616–627.
- Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, 60, 685–700.
- (1970). Mental events. In L. Foster & J.W. Swanson (eds.). *Experience and Theory*. London: Duckworth.
- Dennett, D.C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.
- Fischer, J.M., & Ravizza, M. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Glimcher, P.W., Fehr, E., Camerer C., & Poldrack, R.A. (Eds.). (2009). *Neuroeconomics: Decision Making and the Brain*. London: Elsevier Academic Press.
- Gollwitzer, P.M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, 38, 69–119.
- Gomes, G. (2002). Problems in the timing of conscious experience. *Consciousness and Cognition*, 11, 191–197.
- Glynn, I.M. (1990). Consciousness and time. *Nature*, 348, 477–479.
- Haggard, P. (2011). Does brain science change our view of free will? In R. Swinburne (ed.). *Free will and Modern Science*. Oxford: Oxford University Press.
- Haggard, P. & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, 126, 128 – 133.
- Haggard, P. & Libet, B. (2001). Conscious intention and brain activity, *Journal of Consciousness Studies*, 8, 47-63.
- Haynes, J.D. (2011). Beyond Libet: Long-term prediction of free choices from neuroimaging signals. In W. Sinnott-Armstrong & L. Nadel (eds.), pp. 85–96.
- Hobart, R. E. (1934). Free will as involving determinism and inconceivable without it. *Mind*, 3, 1–27.
- Joordens, S., van Duijn, M., & Spalek, T.M. (2002). When timing the mind one should also mind the timing: Bias in the measurement of voluntary actions. *Consciousness and Cognition*, 11, 231–240.
- Kane, R. (1996). *The Significance of Free Will*. Oxford: Oxford University Press.

- Keller, I., & Heckhausen, H. (1990). Readiness potentials preceding spontaneous motor acts: Voluntary vs. involuntary control. *Electroencephalography and Clinical Neurophysiology*, 76, 351–61.
- Libet, B., Gleason, C.A., Wright, E.W., & Pearl, D.K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain*, 106, 623–42.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Science*, 8, 529–66.
- (1999). Do we have free will? *Journal of Consciousness Studies*, 6, 47–57.
- (2002). The timing of mental events: Libet's experimental findings and their implications. *Consciousness and Cognition*, 11, 291–299.
- Malle, B.F. (2004). *How the Mind Explains Behavior: Folk Explanations, Meaning, and Social Interaction*, Cambridge, MA: MIT Press.
- (2011). Time to give up the dogmas of attribution: An alternative theory of behavior explanation. *Advances in Experimental Social Psychology*, 44, 297–352.
- Mele, A.R. (2009). *Effective Intentions*. Oxford: Oxford University Press.
- Miller J., Shepherdson P., & Trevena J. (2011). Effects of clock monitoring on electroencephalographic activity: Is unconscious movement initiation an artifact of the clock? *Psychological Science*, 22, 103–109.
- Parfit, D. (1997). Reason and motivation. *Proceedings of the Aristotelian Society*, Supplementary Volume 71, 99–130.
- Pockett, S., Banks, W.P., & Gallagher, S. (Eds.). (2006). *Does Consciousness Cause Behavior?* Cambridge, MA: MIT Press.
- Schlosser, M. E. (2012a). Causally efficacious intentions and the sense of agency: In defense of real mental causation. *Journal of Theoretical and Philosophical Psychology*, 32, 135–160.
- Schlosser, M. E. (2012b). Free will and the unconscious precursors of choice. *Philosophical Psychology*, 25, 365–384.
- Schurger, A., Sitta, J.D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *PNAS*, 109, 42, E2904–E2913.
- Sinnott-Armstrong, W. & Nadel, L. (Eds.). (2011). *Conscious Will and Responsibility: A Tribute to Benjamin Libet*. New York: Oxford University Press.
- Schlegel, A., Alexander, P., Sinnott-Armstrong, W., Roskies, A., Tse, P. & Wheatley, T. (forthcoming). Barking up the wrong free: Readiness potentials reflect processes independent of conscious will. *Experimental Brain Research*.
- Soon, C.S., Brass, M., Heinze H.J., & Haynes, J.D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience*, 11, 543–545.
- Van Inwagen, P. (1989). When is the will free? *Philosophical Perspectives*, 3, 399–422.
- Vihvelin, K. (2004). Free will demystified: A dispositional account, *Philosophical Topics*, 32, 427–50.
- Webb, T.L., & Sheeran, P. (2006). Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. *Psychological Bulletin*, 132 (2), 249–268.
- Wegner, D.M. (2002). *The illusion of conscious will*. Cambridge: MIT Press.
- Wolf, S. (1990). *Freedom within reason*. Oxford: Oxford University Press.
- Zhu, J. (2003). Reclaiming Volition: An alternative interpretation of Libet's experiment. *Journal of Consciousness Studies*, 10, 61–77.