Why Libet-Style Experiments Cannot Refute All Forms of Libertarianism

László Bernáth

Introduction

Since Benjamin Libet published the results of his well-known experiments (Libet 1985; Libet–Gleason–Wright–Pearl 1983), it has been heavily debated whether these results refute the existence of free will. Most philosophers who are experts on the topic of free will have reached the conclusion that Libet's original experiments and other Libet-style experiments have not provided enough evidence for denying free will yet.¹ However, the problem as to whether Libet-style experiments could in principle refute free will has not been discussed as much and it seems that there is no consensus on this matter.

Recently, Marcelo Fischborn (2016, 2017) has attempted to shed light on why Libet-style experiments can in principle refute libertarian theories of free will. According to Fischborn, Libet-style experiments can in principle refute libertarian free will because (i) libertarian free will is incompatible with a local determinism in the brain that would make choice predetermined by unconscious brain states and (ii) Libet-style experiments are in principle able to support that there is such a local determinism in the brain.

Against Fischborn, Adina Roskies and Eddy Nahmias (2017) have argued in accordance with their earlier papers (Roskies 2006, Nahmias 2014) that Fischborn is wrong because it is not true either that libertarian free will is incompatible with local determinism or that Libet-style experiments are able to support local or universal determinism.

Although I think that this debate merits attention, both sides share a false presupposition, namely, that the different libertarian theories are similar to each other with regard to what they claim about the role and location of indeterminism in free decisions. Fischborn seems to think that libertarians agree

¹ The fact that the philosophical debate about free will is flourishing without discussion of Libet-style experiments proves this point in itself. Nevertheless, there are many works which explicitly state this (e.g. Mele 2009, 2011; Walter 2011, Shields 2014).

(or should agree) that *decisions about actions*, between which there are no freedom-relevant metaphysical differences, have to be undetermined by previous (mental) states and events in order to the agent has libertarian free will. Roskies and Nahmias argue that only universal (and not local) determinism is incompatible with libertarian free will. Both presuppositions about libertarianism are false and this is why neither Fischborn nor Roskies and Nahmias see that different libertarian theories are vulnerable to the Libet-style experiments to various extents.

In my paper, I spell out which types of libertarian theories can be refuted by Libet-style experiments and which cannot. I claim that, on the one hand, some forms of deliberative libertarianism and restrictive libertarianism cannot even in principle be denied on the basis of these experiments; and on the other hand, standard libertarianism, along with some versions of restrictive and deliberative libertarianism, can in principle be refuted by these experiments. However, any form of restrictive libertarianism can be refuted in the future only if researchers perform new and “untraditional” Libet-style experiments. This is because “traditional” Libet-style experiments investigate decisions in Buridan-type situations. But these decisions are irrelevant with regard to free will, according to the restrictivists.

In the first section, I clarify some terminological issues in order to set the stage for a precise analysis. In the second, I attempt to show what the main reason is for thinking that Libet-style experiments *seem to be* problematic for libertarian theories. Although showing that actions are unconsciously initiated does not pose a real challenge for the elaborated philosophical theories of free will, Libet-style experiments may be problematic for most libertarian theories because, *pace* Roskies and Nahmias, libertarians have good reasons for saying that local determinism and free will are incompatible. In the third section, I argue that because some versions of deliberative libertarianism do not claim that decisions about actions are the center of freedom-relevant indeterminism, Libet-style experiments are unable to refute them after all. In the fourth, I attempt to show why standard centered libertarianism according to which there are no relevant metaphysical differences between different types of conscious decisions is in principle vulnerable to Libet-style experiments to a greater extent. Fifth, I point out the reasons why restrictivist centered libertarianism is less vulnerable to these experiments. Still, I argue that moderate-restrictivism can in principle be denied by Libet-style experiments if these experiments are modified in an important respect. However, proponents of hard-restrictivism should not worry about the possible results of these experiments because these theories restrict the set of free decisions to such a great extent that they cannot be subjected to a proper empirical test.

1 Key Notions of the Debate: “Libertarianism”, “Libet-Style Experiments”, “In Principle”

In Libet’s original experiment (Libet et al. 1983), the subjects are flexed their wrists whenever they wanted. Before they flexed their wrists, they had to memorize the time of their conscious decision to flex when looking at a clock. After flexing their hands, they had to report the time of their initial awareness of their decision or urge to flex their wrists. Libet used Electroencephalography (EEG) to observe electrical activity of the subjects’ brains. He found the following. Insofar as the subjects decided spontaneously, the unconscious occurrence of the so-called Readiness Potential (RP) preceded what subject reported to be the time of their initial awareness of their decision or urge by 350 ms. If the subjects did not spontaneously decide, the time gap was even longer. Libet drew the conclusion that the subjects’ brain unconsciously initiated their actions, 350 ms earlier than they were aware of their decision. Since, in Libet’s view, free will is tied to consciousness, he thought that the subjects did not freely initiate the flexing of their wrists.²

Although some philosophers, neuroscientists, and psychologists embrace Libet’s conclusions (e.g. Wegner 2002, Hallett 2007),³ many of them call Libet’s conclusions and the adequacy of his whole experimental design into question (Roskies 2011, Walter 2011, and Shields 2014 give useful and brief summaries about the methodological issues). Besides measuring and methodological issues, they worry about whether the electrical signal which was measured by Libet is identifiable with unconscious initiations of actions or whether it is something else. It is not clear even more than thirty years after Libet’s first experiments what the RP is besides that it is an electrical activity in the motor cortex that precedes voluntary muscle bursts.⁴ Moreover, some call into question whether the measured RP is connected to voluntary action. There is

2 However, Libet claimed that agents have free will because they are able to stop and not perform the initiated action due to so-called “free won’t” (Libet 1985).

3 Although Daniel Wegner and Mark Hallett agree with Libet that agents do not freely and consciously initiate their spontaneous actions, they do not believe that ‘free won’t’ or something similar exists, which could be the basis of independent conscious control of actions.

4 The two most influential interpretations of RP are rather different. The first interpretation follows Libet’s interpretation according to which RP reflects neural processes associated with unconscious motor preparation of voluntary action. That is, if the RP is earlier than conscious awareness of decision, the initiation of the action is unconscious (Libet 1983, 2004). This interpretation often goes hand in hand with the claim that RP reflects unconscious decision for performing a specific action. The other interpretation claims that RP reflects stochastic fluctuations in neural activity that lead to action following a threshold crossing when humans arbitrarily decide to move (Schurger et al. 2012). In other words, RP reflects such a

some experimental evidence which suggests that the measured RP is a result of Libet's experimental design (Trevena and Miller 2010, Miller et al. 2011).

In order to evade the methodological and measuring issues, many followers of Libet have created more sophisticated versions of the original experiment (Fried et al. 2011). Others, most notably Chun Siong Soon and his colleagues (Soon et al. 2008, 2013), have made more radical changes besides fixing these measuring and methodological problems. Soon and his colleagues did not use EEG. Instead, they focused on observing specific brain regions using fMRI. Since the RP is defined in such a way that it can be detected only using EEG, using fMRI means that Soon's experiment does not attempt to find a correlation between RP and the conscious initiation of action. Moreover, they investigated not only choices about bodily motions but about mental actions as well, given that their later experiment examined decisions about adding or subtracting numbers. They found that the outcomes of choosing between pushing a right-sided or a left-sided button were predictable on the basis of patterns of activation in specific brain regions with roughly 60% accuracy, and 7 seconds before the conscious experience of the choice (Soon 2008). They had similar results in the other experiment with regard to abstract spontaneous choices apart from that they could predict with 60% accuracy only 4 seconds before the conscious awareness of choices (Soon 2013).

Even though Libet's early experiments and Soon's experimental design are very different, both of them are called Libet-style experiments. Why should we put them into the same category? Because I do not think that one should call all neuropsychological experiments about free will Libet-style experiments, and since I did not find any definition of what makes an experiment to be a Libet-style experiment, I propose the following list of the necessary (and together sufficient) conditions for being a Libet-style experiment.

- i. By means of neuroscientific methods, Libet-style experiments try to identify such unconscious brain activities/states that determine (or which are identical with) decisions about actions through identifying correlations between brain activities/states and outcomes of decisions.
- ii. Libet-style experiments examine decisions that are considered to be free decisions by the majority of laymen and/or philosophers.
- iii. In order to ensure that the experiment will investigate decisions that are considered free, the experiment examines decisions in decision-situations

neural mechanism that helps us to make decisions in situations in which there is no good reason for choosing one alternative over another.

for which the influence of unknown intentions and preferences can reasonably be excluded.

- iv. During Libet-style experiments, subjects have to repeat the same type of decision many times.
- v. The researchers have to identify the time when the subjects were aware of making their decision.

Although most of these conditions on being a Libet-style experiment are clear without further explanation, it is worthwhile focusing on (iii.) a little bit more. The reason why the influence of unidentified or unknown prior intentions and preferences should be excluded somehow is that, in many cases, prior conscious states and intentions explain the exact outcome of the decision. Let us suppose that someone consciously decided that whenever she has to decide between two identical buttons, she will push the one on the left. After a while, she faces this “dilemma”, and she instantly pushes the left button without conscious deliberation. Even though this freely-formed intention was the main cause of her action, pushing the button was such a fast reaction that it initiated action unconsciously. Even if a neuroscientist could detect that one of her brain states which was unconscious at the time of the initiation of action determined whether she pushed the right or the left button, many laymen, compatibilist, and libertarian philosophers would think that it does not prove that her decision was not free, because it is probable that the unconscious brain state in question is identical with (or is the neural basis of) her previously freely-formed intention to press the left button. In order to evade these kinds of interpretative possibilities, the researchers have to choose such decision-situations in which they are able to exclude the influence of such distal (and free) preferences or intentions. Without focusing their attention on such a decision-situation, they cannot be sure whether, in some sense, a prior consciously formed intention initiated the action instead of an unconscious one.

This is the reason why Libet and most of his followers examine decision-situations in which there are no objective reasons to prefer one alternative over another one. In Libet’s original experiments, the subjects have to decide when to flex their wrists without any good reason to flex their wrists at any particular time during the whole experiment. In Soon’s experiments, the subjects have no objective reason to prefer pushing the left button/adding over pushing the right button/subtracting.

I call these experiments traditional Libet-style experiments and I will claim that these cannot refute many libertarian theories. The reason is the following. Although restricting the focus of the experiment in such a way is a good way

to exclude the influence of prior and unknown intentions and preferences, it has the disadvantage that the researcher can examine only one type of decision. These are Buridan-type decisions in which there is no reason to choose one alternative instead of any other. But there are some good reasons to think that these Buridan-type decisions are so different to other types of decisions that one cannot generalize results about the former decisions to the latter ones.

As I will argue, this issue has great importance with regard to the refutation of libertarian theories. Luckily, there are other possible ways to exclude the influence of unknown and prior intentions and preferences. For example, Maoz and his colleagues (Maoz 2017) asked their subjects about their relevant preferences in order to reasonably exclude the possibility that there are unknown relevant preferences influencing the choices of the subjects during the experiment. Maoz's research is a new form of Libet-style experiments, and I consider it as one of the first instances of an untraditional Libet-style experiment.

The second terminological issue that I should handle is what it means for an experiment to be capable "in principle" of supporting or proving a particular claim. Fischborn has an illuminating example.

It can be true that physics could in principle show that time travel is possible even if time travel is actually physically impossible. Saying that physics can in principle show that time travel is possible (if time travel is actually possible) says nothing about the truth or plausibility of the possibility of time travel; it only says that physics is the right science for an investigation on the possibility of time travel.

FISCHBORN 2017, 2

One can paraphrase this quote regarding Libet-style experiments in the following way. Saying that Libet-style experiments can in principle show that free will does not exist (if free will actually does not exist) says nothing about the truth or plausibility of the inexistence of free will; it only says that Libet-style experiments are the right experiments for an investigation on the (in)existence of free will.

Fischborn's example is extremely useful, because it can help shed light on an important issue. In general, if one asks whether time travel is possible, she means about "time travel", among other things, that *the body* of the traveler has to travel through time in order that the traveler counts as a genuine time traveler. In this case, physics is the right science for an investigation on the possibility of time travel. However, in an unusual case, one can mean by time travel

such a process in which only a body-independent soul travels to the future or the past. In this unusual case, physics is not the right kind of science for an investigation on the possibility of time travel.

Similarly, it depends on what one means by the term 'free will', if we want to know whether Libet-style experiments can in principle support claims that refute its existence. If one accepts a Hobbesian notion of free will according to which one has free will if she is able to act in accordance with her desires, then Libet-style experiments cannot even in principle show that one has no free will. This is because Libet-style experiments can primarily support only such claims which are based on the "correlation" of some unconscious brain states/events and conscious outcomes of decisions. But these correlations do not say anything about whether the agent has a capacity thanks to which she can act in accordance with her desires. (Because even if such brain states determine our decisions, it does not mean that our actions and desires do not fit each other in most of the cases).

Here, my main interest is whether Libet-style experiments are able to refute libertarian theories of free will. Libertarianism is the conjunction of two theses. One is the thesis of incompatibilism according to which universal determinism is incompatible with free will. The second is the free will thesis which says that adult humans without severe psychological disorders have free will. Even though all libertarian theories accept these theses, there are great differences between them. This is partly because they disagree about i) the location of the relevant indeterminism and ii) what the exact metaphysical structures of different types of decisions are.

I claim that, contrary to Fischborn, Nahmias, and Roskies, the differences which are rooted in these disagreements are relevant with regard to whether libertarian theories can be refuted by Libet-style experiments. Traditional forms of Libet-style experiments can in principle refute only those which claim that a) the location of freedom-relevant indeterminism can be found in decisions about actions *and* b) there is no freedom-relevant difference between Buridan-type decisions and other non-coerced decisions. Consequently, they can refute standard libertarianism *and* some deliberative libertarian theories. Untraditional Libet-style experiments, if they are performed, will be able to refute more types of libertarianism. In principle, they can threaten those which claim that α) the location of relevant indeterminism can be found in the decision process about actions β) there is no freedom-relevant difference between hard decisions in which the agent is strongly motivated to perform more than one action. But neither traditional nor untraditional Libet-style experiments can refute those libertarian theories which say either that A) the location of freedom-relevant indeterminism is not necessarily in the decision processes

about actions or that B) there is a freedom-relevant difference between decisions under moral temptation and every other type of decisions. In the following sections, I will argue for these claims in detail.

2 How Are Libet-Style Experiments Able to Get Libertarians into Trouble?

In the literature, there are two models about how Libet-style experiments deny the existence of free will (libertarian or not). According to the first, Libet-style experiments are relevant to the free will debate because they support the claim that we do not consciously initiate our actions. If it is true, then, as the argument goes, it is also true that our conscious considerations do not influence our actions. And this would mean that we do not have free will. (This kind of argumentation, most prominently, can be found in Wegner 2002. Daniel Wegner argues that even intentions have no causal role in producing our actions).

In my view, this argument is problematic for many reasons. For the sake of brevity, I would like to stress only one difficulty (more detailed criticism can be found in Mele 2009). Even if it is true that we do not consciously initiate our actions, this does not necessarily mean that our conscious considerations do not influence our actions. If a soccer player consciously reaches the conclusion that she should try to deflect the ball after the free kick, her conscious decision influences her bodily motion even if, later, she unconsciously initiates her bodily movement which is aimed at deflecting the motion of the ball. In light of this, even if it is an illusion that there is such an activity as conscious initiation of action, it is still quite possible that our conscious considerations have a major role in controlling our actions.

This problem merits attention because the vast majority of free will theories do not consider the conscious initiation of action as a necessary condition on free will and moral responsibility.⁵ Rather, they stress the importance of specific causal or non-causal influence of the agent and her reasons for action. This is why, if we resist the problematic conclusion that Libet-style experiments can show that conscious considerations do not influence our actions, this strategy seems to be harmless after all.

5 Timothy O'Connor (2000, 122) has lamented that elaborated philosophical theories of free will do not clarify the role of consciousness with regard to free actions. As far as I can tell, experts on this topic still tend to neglect this issue. However, there are some exceptions, for instance, Hodgson 2012.

Now, one may object that Libet-style experiments can prove not only that we unconsciously initiate our actions but also that we unconsciously decide what we will do. In traditional Libet-style experiments, the subjects have no reason to prefer one alternative over the others. Thus, the initiation of action cannot be influenced by conscious reasons and previous conscious decisions. Therefore, the unconscious initiation of the action is the very source of the action. Since the very sources of our actions are the decisions about actions, one can say that the unconscious initiation of the action is the decision itself.

There are two problems. First, even if the unconscious initiation is the main source of the action in question, it does not necessarily mean that considering this initiation to be a decision is a good idea. Let us suppose that our universe is deterministic and the Big Bang is the main source of every event of the universe. This surely does not mean that the Big Bang is any kind of decision. One may argue that there is a good reason for considering unconscious initiations to be decisions from an action-theoretical point of view. But, as far as I know, there is no such argument yet. Second, even if the unconscious initiation is the main source of the action in Buridan-type situations, it does not necessarily mean that it is the case even in situations in which the agent has reasons for choosing one of the alternatives. It is still possible that conscious reasons causally influence this initiation even if the initiation is unconscious. And in this case, most theories of free will is not get into trouble.

The second model for how Libet-style experiments can in principle refute theories of free will is spelled out by Marcel Fischborn. According to Fischborn, who focuses on libertarian theories, Libet-style experiments are in principle able to refute the existence of free will because they can support the following kind of laws.

LD1. For any event x , and any subject s , if an x that is a pattern of neural activity of type B occurs in s 's brain, then s will decide to push a given button.

FISCHBORN 2016, 497

Fischborn suggests a particular way of supporting LD1. Let us suppose that Soon and colleagues will find in the future that there is a 100% correlation between some types of unconscious brain states and some outcomes of Buridan-type decisions. In this case, the neuroscientists could reasonably conclude that there are LD1-like psychological laws which imply local determinism.

Moreover, Fischborn claims that if the neuroscientists could prove many instances of LD1, they would be reasonably able to generalize their results.

Thus, they could in principle support the claim that not only LD₁-type laws are true but the following one as well.

DNC. For any subject *s*, any choice *x*, and any course of action *X*, if *s* chooses to do *X*, then there is a previous event *y* of a type *Y* in *s*'s brain, such that whenever an event of type *Y* occurs in someone's brain, then this subject will choose for the course of action *X*.

FISCHBORN 2016, 497

But if DNC was true, it would mean that our decisions are locally determined by unconscious brain states. From this, according to Fischburn, it follows that libertarian theories of free will are refuted because they are incompatible with this kind of local determinism.

Adina Roskies and Eddy Nahmias have argued that this way of refuting libertarian theories does not work (Roskies & Nahmias 2017). To begin with, it is unsure that any Libet-style experiments (or other neuroscientific experiments) will support anything similar to LD₁. This is because it may be the case that there are multiple realizations of mental states, and/or the thesis of extended cognition may be true, and/or the claim that neural activities are complex and chaotic might be true. Insofar as one of these views is on the right track, it is very unlikely that neuroscientific experiments will find laws similar to LD₁, because they imply that the relations between brain states and mental states are too complicated for the existence of universal and deterministic law-like relations between brain states and decisions.

To my mind, Fischborn has answered to this argument in a very plausible way (Fischborn 2017, 199–200). He has pointed out that this objection has nothing to do with the question whether Libet-style experiments can *in principle* prove LD₁. For the sake of argument, Fischborn accepts that if any of the aforementioned hypotheses are true, then Libet-style experiments cannot support LD₁. (I would even say that this is because if one of these suppositions is true, then LD₁ cannot be true). But it does not mean that Libet-style experiments cannot *in principle* support LD₁. If there are psychological laws such as LD₁, Libet-style experiments could support this claim partly because, if LD₁ is true, then neither multiple realizability, extended cognition, nor the thesis that neural networks are complex and chaotic are true. In other words, if LD₁ is true, Libet-style experiments can show that it is the case.

Roskies' and Nahmias' second objection is based on the claim that libertarian free will is incompatible only with universal determinism. Thus, even if DNC could be proven, it would not refute libertarianism. Roskies and Nahmias are aware that some libertarians have explicitly claimed that free will *is*

incompatible with local determinism (they mention Kane 1996, Ekstrom 2000, Balaguer 2012). However, Roskies and Nahmias argue that libertarians do not have good reason for worrying about local determinism because all of their arguments for incompatibilism support only the incompatibility of universal determinism and free will.

In my view, libertarians argue mainly against the compatibility of *universal* determinism and free will because they try to deny all forms of compatibilism regardless of their exact details. But *most* libertarians deny that agents may have free will if all of their decisions are determined by unconscious brain states. Even if they do not explicitly claim this, one can be sure about it because most libertarians think that the *decisions about actions* should be undetermined by preceding brain or/and mental states in order to count as (directly) free.⁶ More precisely, most libertarians claim that the very event that is choosing an action has to be undetermined. In the literature, these libertarians are called centered libertarians, and most of the detailed libertarian theories fall into this group.⁷

Moreover, centered libertarians do not only claim that *momentary free decisions about actions* are the indeterministic sources of free will but they have *good reasons for* embracing this view (*pace* Roskies & Nahmias 2017). In order to grasp why it is the case, one has to see what libertarians should do for dialectical reasons besides refuting compatibilism. First, they have to show how satisfying the libertarian control-conditions of free action provides more control over action than satisfying compatibilist control-conditions. This is because libertarians, by definition, deny that satisfying compatibilist conditions of free action is sufficient for free action. And if universal determinism rules out free action, the reason for it has to be that it undermines the satisfaction of the control condition on free actions. (The epistemic condition, which is the other widely accepted condition for being morally responsible, has nothing to do with the truth or falsity of determinism.) Second, they need to explain also how the outcome of an indeterministic process need not be a matter of luck. They have to solve this problem because if an action is simply a matter of luck, it cannot be controlled by anything. And an uncontrolled action cannot be free (Levy 2011). Third, they should provide some evidence for that such an indeterministic process which is not simply a matter of luck exists.

6 That is, the location of indeterminism has to be found somewhere in the decision-process. As far as I know, there is only one elaborated libertarian theory which does not accept this claim (Ekstrom 2000). In the next section, I will investigate this theory.

7 According to the article about libertarian theories of free will at Stanford Encyclopedia (Clarke & Capes 2017), all agent-causal and non-causal theories are centered accounts. Even most event-causal libertarians have elaborated centered theories.

In order to solve the second issue, the libertarians have to posit not only an indeterministic process but a *controlled* indeterministic process. In order to reply to the first issue, they have to suppose a controlled indeterministic process that is controlled differently with respect to the way that the compatibilist agent controls her action. Most libertarians have reached the conclusion that the moment of deciding is the event which has to be undetermined in order to solve these problems because decisions about actions are considered to be those events that are controlled in the most direct way. So there is a *hope* that if decisions about actions are the key indeterministic events which are not determined by previous brain/mental states, then they are still *highly controlled* indeterministic events. Additionally, some libertarians think that there is good evidence for libertarian free will if the moment of deciding is considered as the key indeterministic free event. They argue that the phenomenological characteristics of these decisions can serve as an evidence for the existence of highly controlled but undetermined events. At the moment of the decision, the choice seems to be both undetermined and controlled from a first person point of view,⁸ and – according to these libertarians – this phenomenological fact is a proper evidence for believing in the existence of sufficiently controlled and undetermined decisions until there are no strong counter-evidences that outweigh this phenomenological consideration.⁹

Thus, from a libertarian perspective, there are good reasons for denying even psychological and local determinism because they have good reasons to claim that the very moment of decisions about actions are those indeterministic and highly controlled events that are the sources of free will. Or at least, these reasons are no worse than other philosophical reasons for embracing other philosophical views. Rejecting the relevance of local determinism with regard to libertarianism without providing arguments against these reasons is insufficient.

Before I address the problem as to whether Libet-style experiments are in principle able to refute centered libertarian theories, I investigate which non-centered or deliberative libertarian theories are vulnerable to Libet-style experiments.

8 However, some doubt that these decisions seems to be undetermined from a first person perspective. See Nahmias et al. 2004, Horgan 2011, 2012.

9 This phenomenological argument for free will can be found in Reid 1788: 36; C.A. Campbell 1957: 168–174; O'Connor 1995: 196–197; Swinburne 2011: 82. Deery 2015 reconstructs the argument with great care, and I borrowed his list about where the argument explicitly appeared.

3 Are Libet-Style Experiments Able to Refute Deliberative Libertarianism?

In the literature, those libertarian theories are called deliberative which claim that the location of freedom-relevant indeterminism is not (or at least not necessarily) at the *moment* of choice. For example, Laura Ekstrom (2000) claims that it is enough for having free will if the agent's preference formation is indeterministic. Surprisingly, as far as I know, Ekstrom is the only devoted libertarian who adopts a deliberative libertarian theory. However, some experts who officially do not endorse libertarianism have outlined other deliberative libertarian approaches. Daniel Dennett (1978) and Alfred Mele (1995 ch.12) suggested two very similar theories for libertarians. According to Dennett, libertarians should claim that the processes determining which considerations occur to one are the heart of libertarian freedom. Similarly, Mele (1995, 2006: 9–14) shows that the libertarian may claim that the key indeterministic processes are those which determine whether a belief (or desire) comes to the agent's mind.

In principle, Mele's and Dennett's suggestion could be refuted by Libet-style experiments. Let us suppose that someone performs an experiment which is very similar to Soon's and his colleagues' experiment. And say she finds that there is a 100% correlation between different patterns of unconscious brain activities which precede the decision by 10 seconds and the different possible outcomes of the decision. Since it is plausible to suppose that the processes that determine which considerations occur to one are not finished in Buridan-type situations 10 seconds before the choice, this experiment would support strongly the claim that these processes are determined, as well as the outcomes of choices, by unconscious brain states.

However, it is not so simple in the case of Ekstrom's deliberative libertarianist view. Ekstrom's claims, in contrast with other libertarian approaches, that not only decisions about actions but also decisions concerning what desirable is can be undetermined and metaphysically free. According to Ekstrom, the latter kind of decision are not necessarily about what would be desirable to do at a given time but can be about what the good is in general (Ekstrom 2000, 106–109).

Prima facie, this version of libertarianism is not as vulnerable to Libet-style experiments since these experiments investigate only decisions about actions. Even if Libet-style experiments show that decisions about actions are determined by unconscious brain states, it is open to Ekstrom to claim that decisions about what is good are relevantly different in this regard because they are undetermined and free. She could argue, for instance, that all decisions about

actions are driven by conscious or unconscious brain states because our most valuable and rational choices are based on what we consider as good or desirable. But decisions about what is good and what kind of things are desirable may not be determined by other choices. This kind of defense of free will cannot be refuted by reference only to Libet-style experiments because they investigate only decisions about actions. One should deny this argumentation on the basis of philosophical insights or empirical data. But the neuropsychological experiments which could test whether decisions about what is good are determined by previous brain states should be so different to Libet-style experiments that it would be misleading to call them “Libet-style experiments”.

If am right, in contrast with Dennett’s and Mele’s deliberative libertarian theories, Ekstrom’s view, even in principle, cannot be refuted by Libet-style experiments. To put it simply, this is because Libet-style experiments do not investigate decisions about what is good (this may be considered to be judgment rather than decision by many philosophers). To generalize the moral of this, Libet-style experiments are not able to deny those libertarian theories according to which it is not (or not only) decisions about actions that are the location of free will.

4 Centered Libertarianism

4.1 *Standard Versions*

Most libertarians claim that decisions as momentary choices about actions should be undetermined to count as (directly) free. The standard version of this approach says that there is no relevant difference between non-pathological decisions with regard to metaphysical freedom. Every decision is directly free, there are no non-pathological decisions which are unfree or only indirectly free because they are the consequences of prior directly free decisions. In other words, every psychologically “normal” decision is *directly* free regardless of the particular type of the decision in question. This version of libertarianism is the most common. All non-causal (Goetz 1988, Ginet 1990, McCann 1998, Lowe 2008, Pink 2004) and agent-causal theories (Chisholm 1966, Taylor 1992, O’Connor 2000, Clarke 2003, Griffith 2007, Steward 2012) fall into this group. (I know about only one exception. It is Swinburne 2011, which defends a restrictive agent-causal theory). Even some event-causal accounts endorse this unrestricted version of libertarianism (Nozick 1981, Hodgson 2012).¹⁰ In this section, I would like to point out that this unrestricted view on free decisions

10 There are accounts which are between restrictivist and non-restrictivist views. They accept that there are indirectly free and determined decisions but they do not say that

has the price that these standard centered libertarian theories are quite vulnerable to Libet-style experiments.

Libet-style experiments can, in principle, show that decisions are predetermined in Buridan-type situations. For instance, if Soon and colleagues were able to predict on the basis of the activity of unconscious brain states whether the agent will push the left or the right button with 100% accuracy 7 seconds before the conscious awareness of choice, then this would provide strong evidence for unconscious determination with regard to Buridan-type decisions. But, and this is the point, insofar as there is no relevant metaphysical difference between non-pathological conscious choices, this result would provide strong evidence that all conscious decisions are determined by unconscious brain states. Without question, if it turned out to be true, it would refute centered libertarianism.

But why do so many libertarians think that there is no relevant metaphysical difference between non-pathological conscious decisions? The reason why is that there is no *phenomenological* difference between conscious decisions with regard to the prior indeterminacy of these decisions. All conscious decisions about actions seem to be undetermined from a first person point of view. At the moment of conscious decision, agents have the impression whether they decide for *A* alternative or *B* alternative is settled only at the very moment of the choice in question. This phenomenological characteristic of decisions is relevant for many libertarians because they regard it as the best evidence for having free (highly controlled but indeterministic) decisions. If libertarians do not claim that all decisions which phenomenologically seem to be free actually are free, then they cannot maintain that phenomenological traits reliably indicate whether a decision is actually free; therefore, they have to look after another justification for believing in free will. And this is not an easy task. Consequently, even if claiming that every non-pathological decision is free has the cost that the theory will *in principle* be vulnerable to Libet-style experiments, it may be worthwhile sticking to this idea. Until standard centered libertarianism is only *in principle* vulnerable to Libet-style experiments and not refuted by them, the proponents of standard centered libertarianism can reasonably hope that their theory will not be refuted even in the future.

However, many who initially regard centered libertarianism as a plausible view on free will would not like to take this risk. After all, this conclusion about standard centered views says that even a version of the well-known Libet-style

specific types of decisions are necessarily determined. For instance, it seems to me that Mele (2006) offers such a libertarian account.

experiments can *in principle* show that free will is only a *phenomenological illusion*. Luckily, for those who would like to neither reject centered libertarianism nor take this risk, there are other versions of centered views that are *in principle* less vulnerable to Libet-style experiments.

4.2 *Restrictivist Versions*

Libertarian restrictivists claim that only some types of our decisions about actions are *directly* free (Campbell 1938/2013, van Inwagen 1989, Balaguer 2004; Kane 1996, 2000, 2007). This is because, in their view, only some types of decisions can be indeterministic and sufficiently controlled at the same time. However, they tend to call “free” even determined or not sufficiently controlled decisions which are properly influenced by *directly* free decisions.

There are, so to say, hard-restrictivists who hold that only one type of decision can be directly free. For instance, C.A. Campbell (1938/2013) famously claimed that only decisions that are made in situations of moral temptation are directly free. Other restrictivists, for instance Peter van Inwagen and Robert Kane, say that there are more types of decisions which are directly free. Kane claims that all decisions are directly free in which either the agent has more than one motivationally viable option or the agent must make efforts to sustain her intentions against desires and other conditions which make it difficult to carry out our purposes once chosen (Kane 1996, 2000, 2007). I call this more permissive kind of view moderate-restrictivism.

These theories agree that decisions are not free in Buridan-type situations. This is because, in Buridan-type situations, agents have no reasons to prefer one possible course of action over another (or over doing nothing at all). For example, in Libet’s original experiments, the agents do not have any reason for prefer flicking their wrists over doing nothing at any time during the experiment. The subjects are not motivated in either way; thus, neither option is motivationally viable for them. At best, only one option is motivationally viable from trial to trial from their perspective, and it is that which is suggested for them by their unconscious processes. So showing that those Buridan-type decisions in which the agent has no reasons to prefer any possible course of action are not free is not in contradiction with restrictivist libertarianism. Consequently, results of traditional Libet-style experiments which investigate Buridan-type decisions are, even in principle, not able to refute libertarian restrictivism.

One may disagree on the basis that restrictive libertarians believe that Buridan-type decisions are indeterministic (van Inwagen 1989). So, the objection could go, even traditional Libet-style experiments are able to refute the views of restrictivist libertarians about decisions.

It is true that traditional Libet-style experiments can in principle refute the *views of restrictivists about Buridan-type decisions* but it does not mean that these experiments are able to refute libertarian restrictivism. For what libertarian restrictivism argues is that Buridan-type decisions are not free. Libertarian restrictivists can justify this claim by saying either that they are not free because the outcome of these decisions are a matter of chance or that they are not free because they are determined. Peter van Inwagen supposes that Buridan-type decisions are based on an “internal coin-tossing”, suggesting that the outcome of these decisions are just as lucky as the outcome of a coin-tossing (van Inwagen 1989, 417). However, in the light of traditional Libet-style experiments, the restrictivist could suggest that RP is neural noise that helps to form intentions through generating an urge to choose something in such cases that the agent has no reason to choose any particular alternative. It is even open for the restrictivist to say that neural noise is determined by the exact state of the brain and that it in turn determines which alternative will be chosen. In this case, the restrictivist should not worry even if the researchers could predict the outcome of the decision with 100% accuracy in Buridan-type situations. The restrictivist could say that the neural noise in question has a deterministic role only in Buridan-type situations because, in other situations, the agent decides on the basis of her reasons.

Although traditional Libet-style experiments are harmless against restrictivism, untraditional Libet-style experiments may get moderate-restrictivism into trouble. Moderate-restrictivists agree that if the agent has to decide between two options which, for the agent, seem to be attractive on the basis of different sets of reasons, she will make a directly free decision. For instance, if one has to decide between donating towards the education of poor children and donating towards the care of sick people, and if one has not previously ranked the sets of reasons which motivate these actions, then the one will make a directly free decision.

Even though this decision could be not be investigated by traditional Libet-style experiments, Maoz and his colleagues performed a Libet-style experiment which could, in principle, show that even these hard decisions are determined (Maoz 2017). First, they instructed the subjects to rate how much they would like to support different non-profit organizations (NPO) with a \$1000 donation on a scale of 1 to 7. After that, the subjects had to decide between different NPOs. In some cases, they had to choose while knowing that their choice would raise the probability that the researcher would in fact give a donation for the chosen NPO. In other cases, they know that their choices will not influence whether the chosen or another NPO will get donations. Moreover, the researchers asked the subjects to make completely arbitrary decisions in these latter cases. The

researchers investigated, among other things, the RP signals during the deliberate and arbitrary decisions.

This experimental setup, in principle, could support the claim that even those decisions that are both difficult and based on reasons are determined by unconscious brain states. Let us suppose that they find type-A RP signs if the subject chooses the NPO which is represented at the left side of the screen, and type-B RP signs if the subject chooses the one on the right side, regardless of how difficult is the choice for the subject. In this case, it would be reasonable to suppose that the unconscious RP determined the decision (or it was the decision itself).¹¹

Nevertheless, even this untraditional Libet-style experiment has its limits. It is easy to see that it cannot refute those libertarian theories according to which it is not decisions about actions but decisions about what is good that are the source of freedom. This experiment is still a Libet-style experiment focusing on decisions about actions. The subjects previously formed their opinion on what things are desirable or good, and this is why they could give a definite answer to the question of how much they wanted to support different NPOs. Moreover, this kind of experiment is not able to refute hard-restrictivist libertarianism. Hard-restrictivists say that only one decision-type is directly free. Namely, decisions which are made during moral temptation. It is open to the hard-restrictivist to claim that these decisions are not determined by unconscious brain-states, even if all other decisions are.

For practical reasons, I do not believe that Libet-style experiments could refute hard-restrictivism (Campbell 1938/2013). The main problem is that it is needed that the subjects do not know about the real aim of the experiment. If the subjects know that their decisions between morally good and bad choices are the focus of the experiment, and there is no significant reward for making the morally problematic choice, then the subjects will not be tempted to make the morally bad choice. This is because the fact that they are perceived (and probably judged) by the researchers puts such a sociological and moral pressure on the shoulders of the subjects so that they will be not tempted to do the bad thing. Moreover, promising a high reward for making the morally bad choice would be so unprofessional or unethical that such an experiment is unlikely to be approved by an ethics committee.

11 The paper was not peer-reviewed when I accessed the article (2017–12–29). Therefore, one should not consider these results so relevant yet. But it may be worth mentioning that Maoz and his colleagues do not find such a pattern of RP that researchers have found in Buridan-type cases. If Maoz's and colleagues' results are valid, they are strong evidence that deliberative decisions are different to Buridan-type decisions from a neuroscientific point of view.

For instance, let us suppose that researchers perform a version of the Milgram-experiment. The subject is invited to a research center and the researchers say to her that they are investigating the psychological effects of punishment on the learning and memory. The subject has to give an amount of electric shock to someone in another room if she gives a wrong answer. In the original experiment, the subject has no idea that the real focus of the experiment is not the effect of punishment but whether the subject will comply with the unethical order of the researcher or not. She does not know that the other person who “receives” the shocks is, in fact, a confederate of the researchers and a skilled actor who pretends that she receives actual shocks (Milgram 1963).

It is an essential feature of the experiment that the subjects do not know that their decisions and actions are the real focus of the experiment. It is easy to see why. If they knew the real research question is whether they will obey unethical orders, they would try to do their best in order to preserve their perception of themselves as morally good people. Not to mention that, in this case, it would be important for them to present themselves as morally good persons to the experimenters. In such circumstances, most people would not be tempted to act badly even if the experimenters order them to do so. They would interpret the whole experiment as a trial in which they have the opportunity to be a moral hero.

This is even more problematic if one considers the fact that the subjects have to repeat the same type of decision many times. Even if the researchers can fool the subject once or twice about the real focus of the experiment, it would be extremely hard to do this twenty or thirty times.

I have no idea how a mixed Libet-Milgram-type experiment could be performed. Of course, I cannot rule out a priori that more creative researchers than I are able to figure out a Libet-type experiment which can in principle refute hard-restrictivism. But I think the issue that I have highlighted is challenging.¹² So challenging that one can justifiably believe that it is probable that even the best researchers cannot construe such a Libet-type experiment which could in principle refute hard-restrictivism. In sum, the main problem is that the researchers have to use some kind of device to scan the brain activity of the

12 As an anonym reviewer of the paper pointed out, hard-restrictivism is so implausible that it may not worth to spend so much time and space to discuss the difficulties of devising an experimental situation to probe it. However, I think the issue of decisions under moral temptation is relevant even if one does not care about hard-restrictivism or libertarianism at all. If someone would like to do neuroscientific experiments about decisions under moral temptations, in the light of my arguments, one should conclude that it would be wise to choose a non-Libet-style design to get some relevant result.

subject, but it seems to be impossible to use such a device which is undetectable by the subjects. And the subjects would be curious about what the device does. The researchers could lie about anything but they could not plausibly deny that the aim of the device is to collect data about the subject.

Conclusion

I argued that traditional Libet-type experiments which investigate decisions in Buridan-type situations can, in principle, refute only standard (that is, non-restrictivist) centered libertarian theories and such deliberative libertarian views which claim that decision processes about actions are the main sources of freedom. Although we have to wait for the results of untraditional Libet-type experiments, they can, in principle, refute moderate-restrictivist centered libertarian theories as well because they investigate subjects in such situations in which they have motivationally relevant reasons for more than one alternative. However, Libet-type experiments are not able to refute all libertarian theories. First, there are some deliberative libertarian theories according to which indeterministic decisions (or, if you like, judgments) about what is good may be the basis of free will. These theories cannot be refuted by Libet-style experiments because the latter focus on decisions about actions. Second, they cannot refute hard-restrictivist theories which say that agents can bring forth directly free decisions only if they are morally tempted. The problem is that if the subjects know that the researchers observe their behavior and decisions, they will not be tempted to do the morally bad action unless the reward for choosing the morally bad action is high. But it would be unprofessional or unethical to perform an experiment that gives a high reward for doing something morally bad.

Acknowledgements

This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences and the Hungarian Scientific Research Fund number K109638. Moreover, I would like to thank for Alfred Mele, Andrew Cameron Sims, and an anonymous reviewer for helpful notes and advices.

References

Balaguer, M. (2012). *Free will as an open scientific question*. Cambridge, MA: MIT Press.

- Balaguer, M. (2004). Coherent, naturalistic, and plausible formulation of libertarian free will. *Noûs*, 38, 379–406.
- Campbell, C.A. (1938/2013). In Defence of free will. In C.A. Campbell: *In defence of free will: with other philosophical essays*. (pp. 35–55) Routledge.
- Campbell, C.A. (1957). *On Selfhood and Godhood: The Gifford Lectures Delivered at the University of St. Andrews During Sessions 1953–54 and 1954–55*. London: Allen & Unwin.
- Chisholm, R.M. (1966). Freedom and action. In K. Lehrer (ed.) *Freedom and determinism*. (pp. 11–44). New York: Random House.
- Clarke, R. (2003). *Libertarian accounts of free will*. New York: Oxford University Press.
- Clarke, R., & Capes, J. (2017). Incompatibilist (Nondeterministic) Theories of Free Will. E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition). URL = <<https://plato.stanford.edu/archives/spr2017/entries/incompatibilism-theories/>>.
- Deery, O. (2015). The Fall From Eden: Why Libertarianism Isn't Justified By Experience. *Australasian Journal of Philosophy*, 93(2), 319–334.
- Dennett, D.C. (1978). On giving libertarians what they say they want. In D.C. Dennett: *Brainstorms: Philosophical Essays on Mind and Psychology*, (pp. 286–299) Montgomery, Vt.: Bradford Books.
- Ekstrom, L. (2003). Free will, chance, and mystery. *Philosophical Studies*, 113(2), 153–180.
- Ekstrom, L. (2000). *Free will: A philosophical study*. Boulder, CO: Westview Press.
- Fischborn, M. (2017). Neuroscience and the possibility of locally determined choices: Reply to Adina Roskies and Eddy Nahmias. *Philosophical Psychology*, 30(1–2), 198–201.
- Fischborn, M. (2016). Libet-style experiments, neuroscience, and libertarian free will. *Philosophical Psychology*, 29, 494–502.
- Fried, I., Mukamel, R., & Kreiman, G. (2011). Internally generated preactivation of single neurons in human medial frontal cortex predicts volition. *Neuron*, 69(3), 548–562.
- Ginet, C. (1990). *On action*. Cambridge: Cambridge University Press.
- Goetz, S.C. (1988). A noncausal theory of agency. *Philosophy and Phenomenological Research*, 49(2), 303–316.
- Griffith, M. (2007). Freedom and trying: Understanding agent causal exertions. *Acta Analytica*, 22, 16–28.
- Hallett, M. (2007). Volitional control of movement: the physiology of free will. *Clinical Neurophysiology*, 118(6), 1179–1192.
- Hodgson, D. (2012). *Rationality+ consciousness= free will*. Oxford: Oxford University Press.
- Horgan, T. (2012). Introspection about Phenomenal Consciousness: Running the Gamut from Infallibility to Impotence. In D. Smithies & D. Stoljar (ed.), *Introspection and Consciousness*, (pp. 405–422). New York: Oxford University Press.

- Horgan, T. (2011). The Phenomenology of Agency and Freedom: Lessons from Introspection and Lessons from Its Limits. *Humana.Mente* 15, 77–97.
- Kane, R. (2007). Libertarianism. In J.M. Fischer, R. Kane, D. Pereboom, M. Vargas: *Four views on free will*. Oxford: Blackwell Publishing. 5–44.
- Kane, R. (2000). The dual regress of free will and the role of alternative possibilities. *Noûs*, 34(s14), 57–79.
- Kane, R. (1996). *The significance of free will*. New York: Oxford University Press.
- Levy, N. (2011). *Hard luck: How luck undermines free will and moral responsibility*. New York: Oxford University Press.
- Libet, B. (2004). *Mind Time. The Temporal Factor in Consciousness*. Cambridge, MA: Harvard University Press.
- Libet, B. (1985). Unconscious cerebral initiative and the role of the conscious will in voluntary action. *The Behavioral and Brain Sciences*, 8, 519–566.
- 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, 623–642.
- Maoz, U., Yaffe, G., Koch, C., & Mudrik, L. (2017). Neural precursors of decisions that matter—an ERP study of deliberate and arbitrary choice. *bioRxiv*, 097626.
- McCann, H. (1998). *The works of agency: On human action, will, and freedom*. Ithaca: Cornell University Press.
- Mele, A.R. (2011). Free Will and Science. In R. Kane (ed.), *Oxford Handbook of Free Will, 2nd edition*. Oxford: Oxford University Press.
- Mele, A.R. (2009). *Effective intentions: The power of conscious will*. New York, Oxford University Press.
- Mele, A.R. (2006). *Free Will and Luck*. New York: Oxford University Press.
- Mele, A.R. (1995). *Autonomous Agents: From Self Control to Autonomy*. New York: Oxford University Press.
- Milgram, S. (1963). Behavioral Study of obedience. *The Journal of abnormal and social psychology*, 67(4), 371–378.
- 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(1), 103–109.
- Nahmias, E. (2014). Is free will an illusion? Confronting challenges from the modern mind sciences. In W. Sinnott-Armstrong (ed.), *Moral psychology, Vol. 4: Free will and moral responsibility*, (pp. 1–25). Cambridge, MA: MIT Press.
- Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2004). The Phenomenology of Free Will. *Journal of Consciousness Studies*, 11, 162–179.
- Nozick, R. (1981). *Philosophical Explanations*. Cambridge: Mass. Belknap Press.
- O'Connor, T. (2000). *Persons and causes: The metaphysics of free will*. Oxford: Oxford University Press.

- O'Connor, T. (1995). Agent Causation. In T. O'Connor (ed.), *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, (pp. 173–200). New York: Oxford University Press.
- Pink, T. (2004). *Free will: A very short introduction* (Vol. 110). Oxford: Oxford University Press.
- Reid, T. (1788). *Essays on the Active Powers of Man*. Edinburgh, Bell & Robinson.
- Roskies, A.L. (2011). Why Libet's studies don't pose a threat to free will. In W. Sinnott-Armstrong (ed.), *Conscious will and responsibility*, (pp. 11–22). New York: Oxford University Press.
- Roskies, A. (2006). Neuroscientific challenges to free will and responsibility. *Trends in cognitive sciences*, 10(9), 419–423.
- Roskies, A., & Nahmias, E. (2017). "Local determination", even if we could find it, does not challenge free will: Commentary on Marcelo Fischborn. *Philosophical Psychology*, 30(1–2), 185–197.
- Shields, G.S. (2014). Neuroscience and conscious causation: Has neuroscience shown that we cannot control our own actions?. *Review of Philosophy and Psychology*, 5(4), 565–582.
- Schurger, A., Sitt, J.D., & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences*, 109(42), E2904–E2913.
- 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), 543–545.
- Soon, C.S., He, A.H., Bode, S., & Haynes, J.D. (2013). Predicting free choices for abstract intentions. *Proceedings of the National Academy of Sciences*, 110(15), 6217–6222.
- Steward, H. (2012). *A metaphysics for freedom*. Oxford, Oxford University Press.
- Swinburne, R. (2011). Dualism and the determination of action. In R. Swinburne (ed.), *Free will and modern science*, (pp. 63–83). Oxford: Oxford University Press.
- Taylor, Richard (1992). *Metaphysics* (4. ed.), Englewood-Cliffs, Prentice-Hall.
- Trevena, J., & Miller, J. (2010). Brain preparation before a voluntary action: Evidence against unconscious movement initiation. *Consciousness and cognition*, 19(1), 447–456.
- Van Inwagen, P. (1989). When is the will free?. *Philosophical Perspectives*, 3, 399–422.
- Walter, H. (2011). Contributions of Neuroscience to the Free Will Debate: From random movement to intelligible action. In R. Kane (ed.), *Oxford Handbook of Free Will*, 2nd edition. Oxford: Oxford University Press.
- Wegner, D. (2002). *The illusion of the conscious will*. Cambridge, Mass: MIT Press.