

## Surviving Libet: Why Neuroscience Did Not Disprove Free Will (Yet)

**Joachim Nicolodi**

King's College London

### **ABSTRACT**

The idea that we have free will in at least some of the decisions that we make is under heavy attack by recent neuroscientific studies which largely go back to a ground-breaking article published by Benjamin Libet and colleagues in 1983. Libet discovered a specific electrical charge in the cortex, the so-called *readiness potential*, that can be measured significantly before the conscious decision to act arises. This, Libet thought, is the definite proof that it is not *us* that consciously initiates our decisions but unconscious brain activity, thus definitely disproving the existence of free will in any ordinary sense of the term. This paper claims that Libet was wrong to draw such a conclusion from his research. I will explain how his experiments were conducted and then consider two ways one might challenge his conclusion. One popular way of doing this has been unsuccessful, while another gives us good grounds to refute Libet's argument. It is plausible to say that the experiments conducted by Libet, and all Libet-style experiments that came after, capture but a very small percentage of the decisions that we make, and it is therefore premature to draw such stark conclusions from them. I will suggest a new methodological approach to solve this problem and hopefully move the whole debate in a more fruitful direction. Rather than racking our brains about Libet-style experiments and developing more sophisticated approaches to escape or accommodate their empirical observations, we should wait for research to be published that captures the complexity of decision-making more adequately.

### **KEYWORDS**

Free Will, Determinism, Philosophy of Psychology, Neuroscience, Benjamin Libet

## INTRODUCTION

Ordinarily, we think that at least some of our decisions are the product of free will. This notion is dismissed as illusionary by most contemporary psychologists and neuroscientists:

It seems we are agents. It seems we cause what we do. [...] It is sobering and ultimately accurate to call this all an illusion. (Wegner 2002, 341-342)

The phenomenological feeling of free will is very real [...] but this strong feeling is an illusion, just as much as we experience the sun moving through the sky, when in fact it is we who are doing the moving. (Bargh 2008, 148-149)

Free will *is* an illusion. Our wills are simply not of our own making. [...] Your brain has already determined what you will do. (Harris 2012, 5, 9)

This scepticism largely traces back to a ground-breaking study published by Benjamin Libet and colleagues in 1983. This paper aims to show that Libet – and indeed all Libet-style experiments that came after – do not succeed in showing that free will is an illusion and our intuitions about conscious decision-making are misguided.

Here is an overview of what follows. In part one, I will present a definition of free will and then a summary of Libet's study and the philosophical argument he based on it. Free will can be understood in a plethora of different ways, so we do well in getting clear on what we mean when discussing it in this paper. In the second and third parts of the essay, I will discuss two lines of attack that can be brought up against Libet. While the first one fails, the second gives us very good reasons to doubt Libet's conclusions on the inexistence of free will: the experiments from which those conclusions are drawn simply do not capture all the decisions that we care about when debating free will. I will conclude by proposing a novel methodology that might enable neuroscience to answer the question once and for all.

## I. THE TARGET AND THE CHARGE

Most people believe that we have free will in at least some of the decisions that we make (Nahmias et al. 2005). What exactly free will amounts to is a much-debated question within philosophy and trying to come up with an adequate definition would lead us beyond the scope of this paper. For our purposes, it will suffice if we stick to the definition of free will that most people associate with the term, which goes roughly like this:

A subject  $S$  freely wills an action  $\varphi$  at time  $t$  if and only if it was possible, holding fixed everything up to  $t$ , that  $S$  choose or do otherwise than  $\varphi$  at  $t$ .<sup>1</sup>

Let us unpack and elaborate on this in the following. Say I am confronted with a decision and have to choose between option  $a$ , going out, and option  $b$ , staying home to study. I freely will action  $a$  if and only if I could have also chosen  $b$ . If I did not have that possibility, my decision would have been coerced or at least predetermined, thus leaving no room for my conscious volition to intervene in any way. But if the possibility is given, the decision is “up to me” and thus free because it was initiated by my consciousness that could have also pulled in another direction.

Now, of course it is in principle always possible to end up with a different decision, simply because a different past could have led to a different outcome. In one possible universe, I forgot that I have an assignment tomorrow, thus ending up choosing  $b$  over  $a$ . To avoid that our decisions are always free because they always could have turned out differently, we need to add the caveat of “holding fixed everything up to  $t$ ”: we only consider the actual world and the decision made therein, not the outcomes in all possible universes.

To sum up: I act according to my free will if and only if I could have chosen otherwise, that is to say, if I could have consciously decided to do otherwise than I did. Traditionally, also our idea of moral responsibility is based on this fundamental freedom, for could an agent really be blamed for his actions if they were determined by forces completely outside of his influence? I will run with this definition of free will and the connection to consciousness throughout the paper

---

1. I follow here roughly what Timothy O'Connor calls the “Categorical Analysis” of free will (O'Connor 2022).

since Libet himself adopts a similar one in numerous places (Haggard and Libet 2001) and uses “free will” and “conscious will” interchangeably (Libet 2005). It is also the definition that is prevalent in folk psychology, and the one that most psychologists and neuroscientists who comment on the issue associate with the term. Most of these psychologists and neuroscientists believe that Libet’s study conclusively proves that this traditional notion of free will is an illusion and indeed Libet himself thinks that free will does not hold up to neuroscientific scrutiny. But how exactly does his study put the notion of free will in dire straits?

In the 1983 study conducted by Libet and colleagues, participants had to perform a simple physical task that would, according to Libet, classify as a typical free voluntary act. Whenever they felt like it, participants had to flex their wrists and, at the same time, tell when exactly the conscious decision to do so arose. For that purpose, an oscilloscope clock was placed before the participants with a spot of light revolving around the periphery of the screen so that the subjects could tell where the spot was when they were first aware of their intention to act. During this task, the neuronal activity was measured with an electroencephalogram (EEG), and the findings were indeed striking. The researchers found a specific electrical charge in the mesial motor area of the cortex (“readiness potential”, RP) that begins on average 550 msec before the act but also 350-400 msec before the conscious decision (Libet et al. 1983).<sup>2</sup>

We can already see how this is problematic for the definition of free will given above. The *brain* determines the flexing of the wrist, and not the individual’s conscious decision. It is an unconscious neuronal process that initiates our actions and not our consciousness, which is also why we cannot deliberately choose to do otherwise: whatever we do, there is always a RP outside our conscious control that determines which course of action we are going to take. Libet concludes that our traditional view of how we make decisions is fundamentally misguided, as is our way of assigning blame and praise. He grants that our volitions still have a role to play, however only in vetoing our actions while the RP is already present and never in initiating them (Libet 2005). Although Libet leaves this possibility open and wants to grant at least some sort of conscious freedom, we can still say with confidence that he radically contradicts the typical notion of free will and leaves

---

2. It should be noted that the term “readiness potential” was not coined by Libet but goes back to the German “Bereitschaftspotential” in a paper published by Kornhuber and Deecke in 1965. Also the experiment itself can be seen as a precursor for Libet’s, which however gained much more fame.

us only with “free won’t” (Haggard and Libet 2001, 48). To get a clearer picture of the debate and the approaches that different thinkers took, we can summarise Libet’s argument as follows (although he never formalises it this way):

1. We have free will if and only if at least some of our decisions are made consciously.
2. Neuroscience shows that no decision is made consciously.

---

C. We do not have free will.

The argument is logically valid, as can be readily seen. We now have to determine if it is sound. Most of the approaches in the literature can be summarised as either denying the first or the second premise. To give an adequate picture of the debate and come up with a satisfying conclusion of whether or not Libet is successful, I will consider these approaches in turn in the following two sections.

## **II. ROAD ONE: DENYING THE NEED FOR CONSCIOUSNESS**

The first possibility to attack Libet’s argument is to deny premise one. This road has been taken by David Rosenthal (2002), Neil Levy (2005), and Daniel Dennett (1984). Although their approaches differ, they all deny that we necessarily need to make conscious decisions to have free will. They accept premise two – neuroscience showing that decisions are always the product of subconscious brain processes – but they deny that consciousness and the possibility to do otherwise are necessary for free will. Broadly speaking they all argue, therefore, in compatibilist terms: although Libet proves that there exists a local, neural determinism, this does not exclude the possibility of free will. How do they argue for this?

Rosenthal argues that whether or not our actions are free does not depend on whether they are consciously caused, but on whether they are “fitting comfortably within a conscious picture we have of ourselves and of the kinds of things we characteristically want to do” (Rosenthal 2002, 219). If we reflect on ourselves, we all have a certain picture of who we are and what we want, and actions are free if they align with this self-conception, no matter how they are caused. Rosenthal’s

theory has some obvious flaws, which is why I will not spend much time on it. The most devastating is that it radically contradicts our moral intuitions if we grant, as stated in the beginning, that freedom and moral responsibility go hand in hand. If we rob a shop and knock out the cashier, it would be of no excuse whatsoever to say that it contradicts the conception we have of ourselves and is something we would have never pictured ourselves doing.

More interesting proposals are made by Dennett and Levy, which are sufficiently similar to treat them together.<sup>3</sup> They hold that the first premise of Libet's argument is incoherent: decisions are *necessarily* unconscious, which is why – when Libet says decision-making has to be conscious to count as free – he is making demands that free will cannot even in principle fulfil. The question of consciousness is simply irrelevant to the question of free will, which is why the first premise lacks any basis and Libet's experiments do not affect free voluntary decisions in any way.

Levy asks us to consider an ordinary example of decision-making, namely deciding if we should accept a job offer in a different city (Levy 2005, 71). If we look close enough, Levy says, consciousness does not play the role we normally think it does. We do not *decide* that this particular reason, being close to our family, say, is more important than a higher salary. We just *become aware* of these different values, but consciousness does not have any influence in assigning them. Neither can we consciously determine the outcome of our deliberation; it happens on a sub-personal level (that is, not on the level of the conscious agent), and then just appears before us. Dennett argues in the same vein, although not directly mentioning Libet:

But those same decisions can also be seen to be strangely out of our control. We have to wait and see how we are going to decide something, and when we do decide, our decision bubbles up to consciousness from we know not where. We do not witness it being *made*; we witness its *arrival*. (Dennett 1984, emphasis by the author)

I think this line of reasoning has some initial plausibility. Indeed, there seem to be many cases in which a decision just happens to us and we do not have

---

3. Dennett criticises Libet also on other points, for instance in *Freedom Evolves* (2003). However, I think the point he makes here is particularly interesting.

any conscious control over it. Choosing a romantic partner might be one good example, but also choosing a particular kind of career or university course. When we are in love it is very hard to pinpoint why we chose that person over another, and – when we are asked why exactly we love that person – awkward quandaries are commonplace. The same can be said in seemingly more conscious decisions, for instance, when choosing a career as an artist or choosing to study literature. In these cases, too, it is hard – if not impossible – to track exactly how this decision was made. One can of course start talking about passions or talents, but it still does not seem to be an active decision but rather something that happens on the subconscious level.

However, I do not think Dennett and Levy succeed in showing that decisions are *necessarily* like this and cannot be initiated otherwise. There are cases in which it is quite clear that conscious reasoning was involved in the process. To give just one very striking example it is worth looking at Effective Altruism. The whole point of Effective Altruism is to come up with a value system that enables us to assign each possible action a precise numerical value (so-called “quality-adjusted life years”, or QALYs), so that different actions can be compared and the option with the highest value chosen (MacAskill 2015, 61). Moral questions like these seem to be the exact opposite of decisions as described by Dennett and Levy. Every aspect of the decision is meticulously thought through and quantified to eventually lead to the best possible outcome.

Of course, Dennett and Levy could say that here, too, the value of the different options just *appears* before us, independently of our consciousness. However, this would lead to the extreme conclusion that a great part of moral philosophy, and indeed all of value theory, is utterly useless since it rests on unconscious processes we do not have any influence on. This highly counterintuitive conclusion is reason enough to refute Dennett’s and Levy’s claim about the nature of decision-making, which is why I will not go into their compatibilism any further.<sup>4</sup> It is also reason enough to abandon the first way of refuting Libet’s claims. The more plausible way is to argue against the second premise, as I shall explain in the following section.

---

4. Levy further supports his claim by arguing that influencing our decisions is an attempt to control our control system, and therefore conceptually impossible because it would lead to an infinite regress. However, I think his account of a “control system” needs to be fleshed out in more detail to be a valid counterpoint. Like this it is rather ad hoc and unclear, as Mele remarks as well (Mele 2008, 111).

### III. ROAD TWO: QUESTIONING THE SCOPE OF THE EXPERIMENT

As we have seen, trying to refute the first premise of Libet's argument and unlinking free will from consciousness altogether likely leads to some pitfalls. The more promising approach is to deny premise two: that neuroscientific experiments like Libet's prove that no decision is made consciously. One obvious way to do this is by trying to find flaws in the studies themselves and questioning the underlying methodology. However, such methodological flaws will be hard to find since his research has been backed up by many Libet-style studies in more recent years (e.g. Lau et al. 2007, Banks and Isham 2009). The results seem to be accurate (with some obvious deviations in the exact timing of the RP), which is why I will argue that premise two is too strong *even if* we accept Libet's neuroscientific findings. The most devastating way of doing this, I find, is to argue that it is too rash to generalise the findings of Libet-style studies to *all* decisions that we make.

The tasks in Libet's experiment exhibit at least three features that make such a generalisation highly problematic: (i) The tasks are completely *disinterested*, in the sense that the person has no particular reason to flex the wrist at that point and not another. The same can be said for both Lau et al. and Banks and Isham, where the subjects were told to push a button at a point of their choosing and then, again, report the time of their awareness to do so. All these tasks are comparable to Buridan's famous ass, which has to decide between two equidistant and equally big haystacks and therefore has no reason to choose one over the other. (ii) The tasks are spontaneous physical movements that do not need much conscious deliberation to begin with. (iii) The outcome in the experimental tasks is clear, while we normally do not know how the distal decisions that we make will play out.

In this sense, the disinterested, spontaneous, and proximal tasks in Libet-style experiments represent but a very small number of the decisions we make in our lives and seem to be incomparable to the decisions we normally talk about when debating free will. As we have seen, the traditional notion of free will is closely connected to moral responsibility, which is why we are normally concerned with ethical dilemmas when we talk about free will. Consider Sartre's famous example of a young man who is undecided whether he should join the resistance or stay at home and care for his frail mother. Dilemmas like these are, obviously, entirely different from the tasks in Libet's experiments since the person is interested in the



yet unknown outcome and, therefore, spends much time evaluating the different options. Generalising from Libet's experiments to these kinds of cases would be extremely bold, to say the least.

It is important to note that Libet predicts and addresses this kind of worry, however, he does so only briefly. He says that "it is common in scientific researches to be limited technically to studying a process in a simple system; and then to find that the fundamental behaviour discovered with the simple system does indeed represent a phenomenon that appears or governs in other related and more complicated systems" (Libet 2005, 559). Libet gives the example of Millikan and Fletcher measuring the charge of a single electron in one isolated system, saying that it is valid for electrons in all systems. Although Libet does not flesh out this reply in any more detail, I think he has a point. Generalising from case studies is an essential part of the scientific methodology, indeed without it, science would hardly be possible. As the economist Henry Mintzberg pointedly remarks: "If there is no generalizing beyond the data, no theory. No theory, no insight. And if no insight, why do research?" (Mintzberg 2017, 187).

However, it is key that the domain of the generalisation is sufficiently similar to the findings of the case study. And this is not the case with Libet and the Libet-style experiments conducted so far. As we have seen, there are at least three striking differences between the tasks in the research and more complex, moral decisions: (i) disinterest, (ii) lack of deliberation because of the simple task required, and (iii) proximity and knowledge of the outcome. These differences make a generalisation highly problematic since the domain is exceedingly different.

This clearly shows that the neuroscientific research made so far is insufficient to show that all our decisions are made unconsciously, simply because of the striking differences between the simple, Buridan-type tasks in the experiments and more complex, ethical decisions. And, as we have seen, to deny premise two it suffices to show that there are at least some decisions that might be made consciously. This is not to say that more complex decisions are *necessarily* conscious. It is just to say that we do not yet have the research to make any definite conclusions about it. The second premise of Libet's argument fails, and consequently his argument as a whole. Libet's experiment does not demonstrate that our intuitions about free will are misguided since conscious decision-making might still play a role on some occasions.

I will conclude with a brief suggestion about how non-Libet-style experiments could be conducted in the future to yield answers about whether more complex decisions are unconsciously initiated too. To my mind, the best way of conducting such an experiment would be with typical prisoner's dilemmas, a standard example in game theory. I would suggest something roughly like this: put two participants in separate rooms, construct a fictitious scenario that they can imagine (e.g. the classic example of two captured robbers) and then give them options to choose from plus the possible outcomes, of course not with avoiding prison years as an incentive but maybe with winning a small amount of money. If both cooperate with each other and remain silent, they both win £30. If one of them "confesses" to the police and the other does not, the first wins £60 and the latter 0. And if they both confess they both win £15.

This situation would be sufficiently simple to be studied in laboratory conditions, but it would fulfil the three criteria for more complex decisions mentioned above. We are interested in the outcome, we have to think carefully about the different options and the psychology of the other person, and we do not know how our decision will play out. If the results would show a similar neuronal activity as in typical Libet-style experiments, it would indeed be highly problematic for our intuitions about conscious decision-making. As of now, however, the research gives us neither reason for enthusiasm nor worry about free will.

## CONCLUSION

Let us take stock of what we have achieved in this essay. I started by giving the definition of free will that is most commonly used within psychology and neuroscience but also the public discourse. Going into too much detail would have led beyond the scope of this paper, and also missed the target of the discussion since it is explicitly the traditional notion of free will that Libet and most neuroscientists after him challenged. I then turned to the experiments conducted by Libet and colleagues and critically assessed the argument against the traditional conception of free will that he based on them. Responses to this argument can be roughly grouped into two categories: those denying premise one and those denying premise two. I argued that philosophers pursuing the first route – like Rosenthal, Levy, and Dennett – are not able to successfully refute Libet. The more promising way of doing this, I concluded, is to deny premise two by arguing that

Libet's experiments only prove that a small percentage of our decisions are made unconsciously, not decisions *in toto*. By explaining how slightly more complex decision-making processes could be investigated in laboratory conditions, I hoped to show how neuroscience could give us a definite answer to the matter at hand.

## REFERENCES

- Banks, William P. and Eve A. Isham. 2009. "We Infer Rather Than Perceive the Moment We Decided to Act." *Psychological Science* 20 (1): 17-21.
- Bargh, John A. 2008. "Free will is un-natural." In *Are We Free?: Psychology and Free Will*, edited by John Baer, James C. Kaufman, and Roy F. Baumeister: 128-154. Oxford: Oxford University Press.
- Dennett, Daniel C. 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge: The MIT Press.
- Haggard, Patrick and Benjamin W. Libet. 2001. "Conscious Intention and Brain Activity." *Journal of Consciousness Studies* 8 (11): 47-63.
- Harris, Sam. 2012. *Free Will*. New York: Simon & Schuster.
- Lau, Hakwan C., Robert D. Rogers, and Richard E. Passingham. 2007. "Manipulating the Experienced Onset of Intention after Action Execution." *Journal of Cognitive Neuroscience* 19 (1): 81-90.
- Levy, Neil. 2005. "Libet's Impossible Demand." *Journal of Consciousness Studies* 12 (12): 67-76.
- Libet, Benjamin W. 2005. "Do We Have Free Will?" *The Oxford Handbook of Free Will*, edited by Robert Kane, 551-563. Oxford: Oxford University Press.
- Libet, Benjamin W., Curtis A. Gleason, Elwood W. Wright, and Dennis K. Pearl. 1983. "Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity Readiness-Potential." *Brain* 106 (3): 623-642.
- MacAskill, William. 2015. *Doing Good Better: Effective Altruism and How You Can Make a Difference*. New York: Penguin Press.
- Mele, Alfred R. 2008. "Recent Work on Free Will and Science." *American Philosophical Quarterly* 45 (2): 107-130.

- Mintzberg, Henry. 2017. "Developing theory about the development theory." *Handbook of Middle Management Strategy Process Research*, edited by Steven W. Floyd and Bill Woolridge, 177-196. Cheltenham: Edward Elgar Publishing.
- Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner. 2005. "Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility." *Philosophical Psychology* 18 (5): 561-584.
- O'Connor, Timothy, and Christopher Franklin. "Free Will." In *The Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), edited by Edward N. Zalta.
- Rosenthal, David. 2002. "The Timing of Conscious States." *Consciousness and Cognition* 11 (2): 215-220.
- Wegner, Daniel M. 2002. *The Illusion of Conscious Will*. Cambridge: MIT Press.