On Second Thought, Libet-style Unreflective Intentions May Be Compatible With Free Will*

Nick Byrd

Stevens Institute of Technology, NJ, USA

byrdnick.com/contact

Abstract

Some have argued that our sense of free will is an illusion. And some base this free will skepticism on claims about when we become consciously aware of our intentions. Evidence suggests that unreflective intentions form before we are conscious of them. And that is supposed to challenge our sense of free will. This inference from unreflective intention to free will skepticism may seem intuitive at first. However, upon reflection, this argument seems to entail a magical view of free will. So, insofar as free will does not require magic, unreflective intentions do not necessarily undermine our sense of free will.

Keywords: reflection, intention, free will, neuroscience, philosophy of mind

Published in *Logoi* in 2021 at

https://revistasenlinea.saber.ucab.edu.ve/index.php/logoi/article/view/5025

---

* Acknowledgements: This paper's origin was a blog post that benefitted from anonymous commenters on Reddit's r/philosophy as well as more general comments from John Danaher, Helen De Cruz, Marylin Delgado, Shane Littrell, Alfred Mele, Marie Newhouse, Quentin Ruyant, Neil Sinhababu, Sergio Tenenbaum, and Alex Wellerstein.

Bajo ninguna ilusión: las intenciones irreflexivas no socavan el libre albedrío

Resumen

Algunos han argumentado que nuestro sentido de libre albedrío es una ilusión. Y algunos basan este escepticismo en lo que interpretan como el momento en que nos volvemos conscientes de nuestras intenciones. La evidencia sugiere que las intenciones no reflexivas se forman antes de que seamos conscientes de ellas. Y eso se supone que desafía nuestro sentido del libre albedrío. Esta inferencia puede parecer intuitiva al principio. Sin embargo, este argumento parece implicar una visión mágica del libre albedrío. En la medida en que este no requiere magia, las intenciones no reflexivas no necesariamente socavan nuestro sentido de libre albedrío.

Palabras clave: reflexión, intención, libre albedrío, neurociencia, filosofía de la mente

"If the 'act now' process is initiated unconsciously, then conscious free will is not doing it."

–Benjamin Libet (2001, p. 61)

Suppose you are scrolling through a feed of your favorite online content. You are not necessarily keeping track of time, but you are consciously and deliberately deciding what to look at, read, tap, and swipe. In other words, you are intentionally engaging with some contents of the feed and skipping other parts of the feed. Now suppose that you are doing this while neuroscientists are recording your brain activity. Eventually, the neuroscientists are familiar enough with your brain activity and behavior that they can predict your looking, tapping, and swiping before you are even aware of your intention to look, tap, or swipe. Would these predictions undermine your sense of free will?

Some have argued that our sense of free will is an illusion (e.g., Wegner, 2002). And some base this free will skepticism on claims about unreflective intentions that form before we are conscious of them. This inference from unreflective intention to free will skepticism may seem intuitive at first. If I am free to form my own intentions, then I should be conscious of them as soon as I form them. So if I am conscious of my intentions only after they form, then I seem unable to form my own intentions. I found this kind of argument intuitively plausible at first. However, upon reflection I think that this argument might entail a magical view of free will. Thus, if free will does not require magic, then unreflective intentions do not necessarily threaten free will.

1. Evidence

The opening thought experiment is the 21st century version of real studies conducted by researchers like Benjamin Libet. In Libet-style studies, participants are instructed to do something

with their hand—e.g., raise a finger, tap a button, flex their wrist, etc. They can do this whenever they want. While participants are deciding what and when to move their hand, they are watching a clock-like device so that they can remember the moment that they were aware of their intention to choose. Throughout the study, researchers are recording the activity of participants' hands and brains. After a participant decides to do something with their hand, the researchers ask the participants to point to the location on the clock at which they became aware of their intention to do something with their hand. The researchers end up with data about the timing of brain activity, hand activity, and participants' awareness of their intentions (Figure 1).

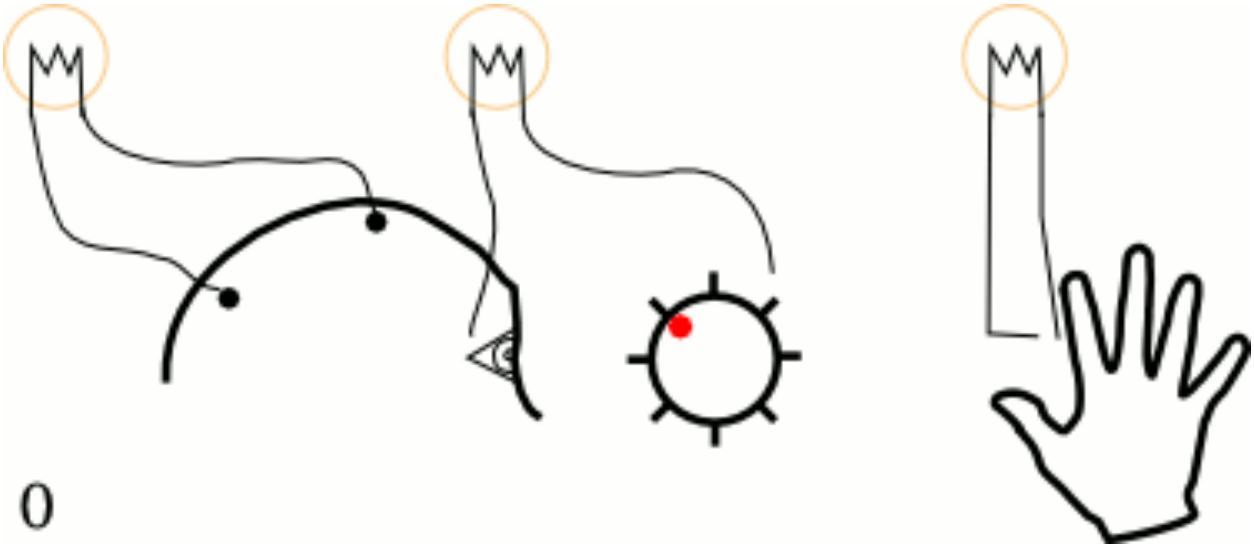

Figure 1. "Benjamin Libet's experiment in which he asked each subject to choose a random moment to flick their wrist while he measured the associated activity in their brain" (Createaccount, 2011).

With these data, researchers have claimed that they can predict whether participants' will move their hands *before* the participant remembered being aware of their intention. The trick is a particular pattern of brain activity that reliably precedes not only participants' hand movement, but

their awareness of an intention to move their hand (Figure 2). Many have found this a provocative result, perhaps because of provocative remarks from Libet himself. “We do not hold people responsible for actions performed unconsciously, without the possibility of conscious control. […] Why then should an act unconsciously developed …over which [we have] no conscious control be regarded as an act of free will?” (B. Libet, 1999, pp. 52–53).

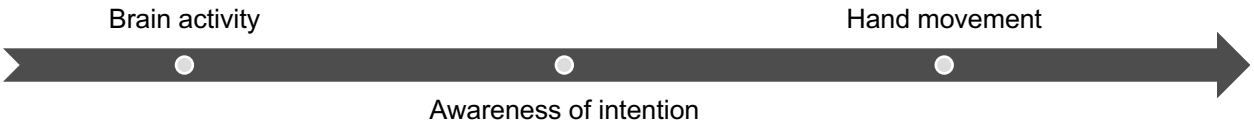

Figure 2. Timeline of events prior to voluntary finger movement, in research by Libet et al. (1983).

Many studies like Libet’s have been conducted. And researchers often claim to be able to predict voluntary actions before people are consciously aware of their intention to act (Koenig-Robert & Pearson, 2019). Some take this evidence about the order of brain activity and intention awareness as grounds for thinking that our sense of conscious ~~and~~ free will is illusory: “the brain started first, followed by the experience of conscious will, and finally followed by action” (Wegner, 2002).

Of course, you may not be convinced. You may reject the scientists’ explications of ‘intention’ (e.g., Mele, 2003), provide an alternative explanation of the data (e.g., Levy & Bayne, 2004), or question the connection between Libet-like protocols and free will more generally (e.g., Saigle et al., 2018). However, suppose you grant the provocative result and its supposed implication: we are not aware of intentions until after intention formation has occurred in our brain. Does that undermine free will? In the past, I thought that it might. However, after more reflection, I am less convinced. In what remains, I will explain my change of mind.

2. Unreflective Intentions

One interpretation of Libet-style results is that people are not consciously aware of their intentions until after their intentions form. Following standard explications of reflective cognition as both deliberate and conscious (Byrd, Revisions Requested; Shea & Frith, 2016), this would mean that intention formation can be unreflective. It might even mean that intentions themselves can be unreflective. This may seem odd to some. As Elisabeth Pacherie tells us, "Many philosophers hold the view that if we do something intentionally, we must be aware of what we are doing. Therefore, they consider that it is of the essence of intentions to be conscious" (Pacherie, 2015, pp. 1-2). However, many other philosophers and cognitive scientists are untroubled by unreflective intentions. Consider just a handful of examples.

- "bad faith involves a conflict between reflective volition (will) and unreflective intention" (Barger, 1976, p. 5).
- "…an understanding of unreflective intentions. Just as one may unreflectively yet intentionally avoid sitting in a position that is uncomfortable (without ever reflecting on that uncomfortable position), so, too, one may unreflectively yet intentionally avoid an uncomfortable venture into reflection (without ever reflecting on the sources of that discomfort)" (Church, 1987, pp. 364-65).
- "The goal-directed behavior specified in an implementation intention is triggered without conscious intent once the critical situational context is encountered" (Gollwitzer, 1999)
- "If many proximal intentions produce actions without the intentions showing up in consciousness, this is not a terribly surprising result" (Mele, 2009, p. 34).

3. Argument

So how might the existence of unconscious intentions challenge some notions of free will? The idea seems to be that if the actual causes of intended behavior are unreflective, then our intended behavior is not free (see Mele, 2013 for a longer discussion). The premises of this kind of *unfree-because-unreflective argument* can take a few forms. Consider one.

1. I am not consciously aware of my intentions as or before they form.
2. If I am not consciously aware of intentions as or before they form, then intentions occur unreflectively. (This follows from our definition of 'reflective')
3. If intentions occur unreflectively, then whatever actions result from the intentions are not free.
4. Therefore, my intentional actions are not free.

These premises may seem intuitively plausible, especially if you share the free will skeptic's intuition about Libet-style cases of unreflective intention. However, consider some of the hitherto hidden features of the unfree-because-unreflective argument.

The free will skeptic's implication: If I am unaware of intentions as or before they form, then whatever actions result from the intention are not free. (This follows from 2 and 3.)

The neuroscience assumption: Intentions can be identified by certain patterns of neural activity.

The temporal observation: The neural correlates of our intentions occur prior to our awareness of our intentions.

The temporal observation is central to the free will skepticism that is motivated by a lack of self-conscious awareness. If we are to reinvigorate free will from the unfree-because-unreflective argument, then this may be the premise to attack. However, I will not dispute the temporal observation. By the end of this paper, it should be clear that free will may be unscathed by the temporal observation.

We might also be suspicious of the neuroscience assumption. For example, we may find ourselves wanting to know precisely why it is intention (as opposed to some other phenomena) that neuroscientists are monitoring during Libet-style studies. However, I cannot resolve that suspicion in the space remaining. Fortunately for the friend of free will, we may not satisfy this curiosity. So I will grant the neuroscience assumption.

This leaves us with the free will skeptic's implication. The point I want to make is that even if we grant the neuroscience assumption and the temporal observation, it is not clear why we should accept the free will skeptic's implication. And if we reject the free will skeptic's implication, then we may be able to reject the unfree-because-unreflective argument. After all, the free will skeptic's implication is a direct consequence of two premises in the unfree-because-unreflective argument.

4. Reframe & Reflection

Consider the third-personal frame of the Libet-style study: a *researcher* can detect our intention before we do. Third-personal prediction of our behavior may be crucial to the intuition that we lack free will (Fischborn, 2016)—especially if this prediction comes from neuroscientists (Nahmias et al., 2007). But how might our intuition change if we reframe Libet-style studies to

focus our own perspective? Rather than ask when a researcher should be able to detect our intention, we can ask when *we* should be able to detect our intention. Should we be aware of our intention before it forms? As soon as it forms? Later?

When I reframe the Libet-style studies first-personally, I find that my initial free will skeptic intuition flips. I become much less concerned about the fact that I may be aware of my intentions only after they form. In case it is not entirely clear why the new frame results in a new intuition, allow me to explain.

4.1 New Frame

We agree that our mental activity corresponds to brain activity—to deny that would be to deny the motivation for cognitive science altogether. So when I form an intention, my brain does something and when I become aware of my intention, my brain does another thing (Figure 3).

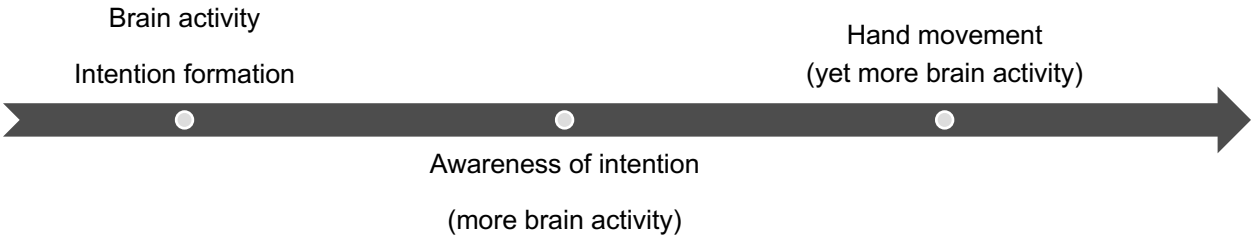

Figure 3. Correspondence of brain activity with intention formation, intention awareness, and subsequent hand movement.

So just as our intentions precede our awareness thereof, the brain activities corresponding to our intentions precede the brain activities corresponding to our awareness of the intention. This is only natural. The neural correlates of my intentions *must* precede my awareness of the intentions and, therefore, the neural correlates of that awareness. My awareness cannot precede its

corresponding brain activity! To suggest otherwise would be to reject the correspondence between mental activity and brain activity that motivates studies like Libet's.

4.2 New Intuition

After reframing the Libet-style studies in terms of my ability to detect my own intentions, my sense of free will no longer feels threatened. It seems utterly unsurprising that my awareness of my intentions occurs after my intention and its corresponding brain activity. Think about the causation involved in free action. Becoming aware of an intention must be causally downstream of—among other things—the existence of the intention. So intention-forming brain activity *must* occur before downstream awareness of intention. How could I have rationally expected otherwise?

So which intuition is better? My prior intuition or my new intuition? One unforgivable problem with my prior intuition is that it seemed to imply some sort of magical thinking. My prior intuition involved a notion of free will according to which I could become aware of my intention before it existed.

5. Conclusion

If my new intuition is right, then one or more premises of the unfree-because-unconscious argument are false, which reinvigorates the possibility that our intentional actions can be free even if the original intention is unreflective at first.

And this makes good sense. The unfolding of Libet-style cases is unsurprising from the first-person sense of free will. After all, when I intend to do something, become aware of this intention, and then do it, the conditions for many notions of free will seem to be satisfied (Dennett, 2015). So if we find that intentions form pre-reflectively, such observations may be entirely

compatible with free will. Indeed, unreflective intentions have been a staple in philosophers' notions of free choice and will both before and after the time of Libet's studies.

- "[For Sartre], unreflective intention is the direct expression of original choice" (Barger, 1976, p. 5).
- "We can see …implicit licensing of the will in our unreflective intentions in that we can and do stop forming them whenever we become aware of or think there are features of our circumstances that may provide us with reasons against these intentions or stronger reason for other competing intentions" (Sanchez, 2011, p. 115).

In fact, it is this latter ability to cancel unreflective intentions that prevented Libet from rejecting free will. "The volitional process is …initiated unconsciously. But the conscious function could still …veto the act. Free will is therefore not excluded" (B. Libet, 1999, p. 47).

References

Barger, Bill. 1976. “Sartre on ‘Original Choice.’” Philosophy Research Archives 2: 1–19. https://doi.org/10.5840/pra197621.

Byrd, Nick. 2020. “All Measures Are Not Created Equal: Reflection Test, Think Aloud, and Process Dissociation Protocols.” Revisions Requested. Thinking & Reasoning.

Caruso, Gregg. 2013. Free Will and Consciousness: A Determinist Account of the Illusion of Free Will. Lexington Books.

Church, Jennifer. 1987. “Reasonable Irrationality.” Mind 96 (383): 354–66. https://www.jstor.org/stable/2254312.

Createaccount. 2011. English: Benjamin Libet’s Experiment in Which He Asked Each Subject to Choose a Random Moment to Flick Their Wrist While He Measured the Associated Activity in Their Brain (in Particular, the Build-up of Electrical Signal Called the Readiness Potential). Own work. https://commons.wikimedia.org/wiki/File:B.Libet.experiment.gif.

Dennett, Daniel C. 2015. Elbow Room: The Varieties of Free Will Worth Wanting. New edition edition. Cambridge: Bradford Books.

Fischborn, Marcelo. 2016. “Libet-Style Experiments, Neuroscience, and Libertarian Free Will.” Philosophical Psychology, March. http://www.tandfonline.com/eprint/KBwhTKyDUHeg3MVphVdG/full.

Gollwitzer, Peter M. 1999. “Implementation Intentions: Strong Effects of Simple Plans.” American Psychologist 54 (7): 493–503. https://doi.org/10.1037/0003-066X.54.7.493.

Koenig-Robert, Roger, and Joel Pearson. 2019. “Decoding the Contents and Strength of Imagery before Volitional Engagement.” Scientific Reports 9 (1): 3504. https://doi.org/10.1038/s41598-019-39813-y.

Levy, Neil, and Tim Bayne. 2004. “Doing without Deliberation: Automatism, Automaticity, and Moral Accountability.” International Review of Psychiatry 16 (3): 209–15. https://doi.org/10.1080/09540260400003909.

Libet, Benjamin. 1999. “Do We Have Free Will?” Journal of Consciousness Studies 6 (8–9): 47–57.

Libet, Benjamin, Curtis Gleason, Elwood Wright, and Dennis Pearl. 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 (3): 623–42.

Libet, Benjamin W. 2001. “Consciousness, Free Action and the Brain: Commentary on John Searle’s Article (with Reply From Searle).” Journal of Consciousness Studies 8 (8): 59–65.

Mele, Alfred R. 2003. Motivation and Agency. Oxford University Press.

———. 2009. Effective Intentions: The Power of Conscious Will. Oxford University Press, USA.

———. 2013. “Free Will and Neuroscience: Revisiting Libet’s Studies.” In , 195–207. Is Science Compatible with Free Will? Springer.

Nahmias, Eddy, D Justin Coates, and Trevor Kvaran. 2007. “Free Will, Moral Responsibility, and Mechanism: Experiments on Folk Intuitions.” Midwest Studies in Philosophy 31 (1): 214–42.

Pacherie, Elisabeth. 2015. “Conscious Intentions.” In Open MIND, edited by Thomas Metzinger and Jennifer M. Windt. Open MIND. Frankfurt am Main: MIND Group. https://doi.org/10.15502/9783958570122.

Saigle, Victoria, Veljko Dubljević, and Eric Racine. 2018. “The Impact of a Landmark Neuroscience Study on Free Will: A Qualitative Analysis of Articles Using Libet and Colleagues’ Methods.” AJOB Neuroscience 9 (1): 29–41. https://doi.org/10.1080/21507740.2018.1425756.

Sanchez, Santiago. 2011. “An Argument for a Reasons-Based Doxastic Voluntarism.”

Setiya, Kieran. 2018. “Intention.” In The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta, Fall 2018. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2018/entries/intention/.

Shea, Nicholas, and Chris D. Frith. 2016. “Dual-Process Theories and Consciousness: The Case for ‘Type Zero’ Cognition.” Neuroscience of Consciousness 2016 (1). https://doi.org/10.1093/nc/niw005.

Wegner, Daniel M. 2002. The Illusion of Conscious Will. MIT press.