Can Neuroscience Provide Conclusive Evidence Against “Free Will”?

Inna Lukyanenko

The University of Melbourne

2018

"Our sense of our own freedom results from our not paying close attention to what it is like to be us. The moment we pay attention, it is possible to see that free will is nowhere to be found, and our experience is perfectly compatible with this truth."

(Sam Harris, "Free Will", p.64)

The felt sense of having free will (or volition), i.e. being free and conscious authors of our thoughts and actions, seems to be fundamental to human experience (Haggard, 2008). The belief in free will has implications not only on personal, but also on the societal level: importantly, it is considered to be the basis for moral responsibility (Nahmias, Coates, & Kvaran, 2007). Indeed, if people are not really accountable for their actions, why would we punish crimes and reward achievements? Perhaps for this reason, the issue still leads to intense disagreements in the scientific community, e.g. Harris vs. Dennett (Harris, 2012; Dennett, 2014) and Bargh vs. Baumeister (Bargh, 2009; Baumeister, 2009). Can neuroscience resolve this long-standing debate? In this essay, I will argue that the philosophical question of the existence of free will is not within the scope of neuroscience. However, I contend that by revealing the mechanisms underlying human intention, neuroscience challenges our traditional understandings of free will (Nahmias, Coates, & Kvaran, 2007; Roskies, 2010). As Harris (2012) points out, our feeling of freedom arises from ignorance about the underlying causes of our actions, and this ignorance is being stripped away by neuroscientific discoveries. As an example of neuroscientific evidence, I will discuss Libet's seminal studies on intention and their implications (Libet, Gleason, Wright, & Pearl, 1983; Libet, 1985). I will then examine some of the critiques of these studies and how they were addressed in replications (Soon, Brass, Heinze, & Haynes, 2008; Haynes, 2011; Soon, He, Bode, & Haynes, 2013). Finally, I will survey some potential ethical implications of these findings.

The main issue when addressing the impact of neuroscience on the question of the existence of free will is that there is no single coherent definition, but rather a highly subjective "gut feeling" described differently by different people (Roskies, 2010). Our intuitive understanding of free will seems to involve two components: the freedom of choice and the idea conscious agency. Both of these, if examined closely, are not straightforward. First, "freedom" implies absence of any internal or external causation, which is incompatible with the scientific understanding of causality. It is hard to imagine how a decision can be truly "free"; on the other hand, there is ample evidence that both internal and external factors (such as previous choices and environmental cues) affect our decisions (Murawski, Harris, Bode, & Egan, 2012; Mattler & Palmer, 2012; Bode et al., 2014). Even if we reject "hard" determinism (every event being completely determined by the prior state of the universe) and accept the possibility of random events, this still does not grant us "freedom" — random actions can hardly be called "free" (Searle, 2001). More fundamentally, experimental science is inherently probabilistic, as one cannot possibly account for every confounding factor; thus, it could never "prove" determinism (Roskies, 2010). Second, consider the idea of conscious agency. We often make a distinction between voluntary actions (that appear to be caused by conscious will) and involuntary responses — there seems to be an assumption that conscious choices are "free", implying that the person could have chosen differently. That is why we judge deliberate actions more harshly than non-deliberate ones (e.g. premeditated murder vs. manslaughter). The very idea of conscious will, however, is incompatible with the scientific assumption of *physicalism*, according to which all mental states (such as conscious thoughts and intentions) are caused by neural processes in the brain. Unfortunately, neuroscience itself is confined to the physical realm and, hence, cannot be used to "prove" physicalism (Batthyany, 2009). Thus, the two fundamental philosophical assumptions required to reject free will (causality and physicalism) are beyond the scope of neuroscience.

While neuroscience cannot resolve the philosophical debate, it can and does challenge our traditional views on free will by revealing the mechanisms underlying intention and behaviour. There are several conceptual aspects of volition, each requiring a different experimental approach: initiation of action, intention, decision-making, inhibition, and subjective experience of volition (Roskies, 2010). I will restrict my focus to studies of intention. A famous series of studies by Libet and colleagues challenged the traditional understanding of the origin of intentions (Libet et al., 1983; Libet, 1985). Participants were asked to flex a finger when they felt the urge to do so and to report the time when the intention entered their conscious awareness. The time was measured by the position of a dot moving around a clock; I will refer to this point in time as the "will". At the same time, an electroencephalogram (EEG) was recorded, and it was found that the so-called readiness potential, known to be associated with movement preparation, began to arise on average 350ms *before* the reported "will". Subsequently, Haggard and Eimer (1999) confirmed and extended Libet's results by using a multiple-choice version of the experiment (choosing to press either the left or the right key). These findings suggest that preparation of movement starts *before* the conscious awareness of the intention, thus, challenging the naïve view that our intentions arise an act of conscious will. Needless to say, many people in science, philosophy and the general public found this implication disturbing and attempted to invalidate it by finding methodological issues with the experiments.

The main criticism of Libet's experiments concerned the reliability of the relative timing of the neural activity and the "will". In order to conclude that neural activity *causes* intention one needs to establish that it *always* precedes "will" (Haynes, 2011). It has been argued that Libet's experiments do not consistently demonstrate this. First, it was suggested that subjective timing judgements of "will" are unreliable (Breitmeyer, 1985). Second, it was noted that, although on average neural activity preceded "will", Libet's data does not provide

evidence that it happened on every single trial, and, in fact, the result could be explained by temporal "smearing" due to averaging across trials (Trevena & Miller, 2002). Subsequently, Libet (2002, 2003) responded to and rejected these criticisms. In any case, potential timing issues were resolved in a replication by Soon et al. (2008), where functional magnetic resonance imaging (fMRI) was used instead of EEG. The clock with a moving dot was replaced by a screen with a stream of letters, and participants were to choose between two response buttons and to report the letter that was on the screen when the decision was made. Then, multivariate pattern classification analysis was used to predict participants' *specific* intentions from the pattern of brain activity — an innovative method that involves classifying patterns of neural activity across multiple voxels in the fMRI data. It was found that predictive information in the brain had already appeared 7 seconds before the conscious decision, with predictive accuracy significantly above the level of chance. In light of this considerable delay (on the order of seconds), the participants' temporal precision of the "will" was less of an issue here than it was in Libet's study. These findings were later replicated by Bode et al. (2011) using ultra-high field fMRI. In the same year, Fried, Mukamel, & Kreiman (2011) used single-neuron recording to show that neural activity in the supplementary motor area was present over 1.5 seconds before the reported decision. However, Batthyany (2009) suggested that the decision studied in these experiments was a simple urge rather than a product of conscious deliberation. This critique was addressed in yet another replication by Soon et al. (2013) that investigated abstract intentions (choosing between addition or subtraction of two numbers on each trial) and led to the same conclusions. The error rate was only 1.2%, indicating that participants indeed performed the mental operations rather than responded randomly. Thus, the evidence about the relative timing of neural events and conscious awareness appears robust.

How does this neuroscientific evidence affect our intuition of being free and conscious agents? Experimental results suggest that our brain "knows" about our intention several seconds before we consciously realise it. If examined closely, this conclusion is consistent with the physicalist view: that all mental events (including conscious intentions) are a result of physical processes. Still, the idea of being "controlled" by our brain has received little enthusiasm; yet, it is not as disquieting as it may seem at the first glance. The aforementioned neuroscientific studies reinforce the idea that unconscious processes are essential to our mental functioning. Consciousness is a recent evolutionary development, and it is limited in capacity and slow to respond, meaning that we are living in a "remembered present" (Edelman, 1989; Edelman, 2003); hence, our survival and functioning largely relies on automatic processes, so-called "zombie systems" (Koch & Crick, 2001; Crick & Koch, 2003). Consciousness itself, its role, and its neural underpinnings are still far from being completely understood. In particular, it is still debated whether it has any adaptive purpose in itself or was developed as a "side-effect" of adaptive advantages such as the ability to have long-term plans and goals (Edelman, 2003). In this context, one can ask a question: does the subjective feeling of free will have any purpose and/or adaptive value?

As discussed above, neuroscientific evidence does not support the idea of free will, so the only evidence that we are left with is our subjective feeling of causality between thought and action. However, this feeling can be erroneous: one can easily "trick" people into inferring causality when there is none — all it requires is time precedence, consistency and lack of another explanation (Wegner, 2003). One can argue that the last condition is *always* fulfilled in the context of our thoughts and actions, as we are unaware of the underlying unconscious processes. Although feelings are very real to the person experiencing them, they do not always reflect the true state of things, particularly if they are based on incorrect beliefs. This leaves us with the question: why do we still believe in free will? There is a

concern that disbelief in free will may have negative societal implications: some studies have found that it increases cheating (Vohs & Schooler, 2008), or reduces helpfulness and increases aggression (Baumeister, Masicampo, & DeWall, 2009). But, as Bargh (2009) points out, the benefits of believing in an illusion may not justify keeping it and are ultimately self-serving: we tend to claim the authorship for behaviours that we are proud of and blame circumstances for those that we are ashamed of. Though it may seem initially distressing and even frightening, rejecting the belief in free will may lead to positive consequences at both the individual and societal level. I agree with Harris (2012) that, once internalised, disbelief in free will only enhances personal experience: it removes anger and increases compassion towards others and, inwardly, it reduces the sense of entitlement and increases appreciation. On the societal level, according to Greene and Cohen (2004), the rejection of free will does not invalidate our justice system, but rather can bring it to a more progressive level by shifting the focus from punishment and retribution to deterrence and rehabilitation. For instance, premeditated murder and manslaughter would still be treated differently, not in terms of culpability, but in terms of predicting the future behaviour of the individual and assessing their potential danger to society (Harris, 2012). This would lead to a more humane treatment of perpetrators of violent crimes, while still protecting society from their actions. Thus, while the awareness of a lack of free will may seem initially disturbing, the benefits are likely to outweigh the negative effects.

To summarise, neuroscientific evidence cannot disprove the existence of free will in the strict philosophical sense; however, it does undermine traditional understandings of free will and conscious agency. Experimental data, though it cannot be taken as conclusive evidence, is consistent with the non-existence of free will: there is no evidence that consciousness can cause intentions, but there is compelling experimental evidence that intentions can be predicted from neural activity preceding the conscious decision. As for the

implications, both the subjective feeling of causality and the fear of negative social consequences may be unwarranted. In fact, there may be benefits in rejecting the traditional idea of free will, such as more humane treatment of criminal offenders. Some important questions still remain to be answered, for instance: how should the modern scientific understandings about the neural mechanisms underlying human behaviour be incorporated within our ethical framework of moral responsibility? Perhaps, it is about recognising and appreciating the privileges that we cannot take credit for, such as intelligence, good character, or social circumstances in which we were born, and using these privileges to improve the experience of those who are less fortunate.

References

Bargh, J. A. (2009, June 23). The will is caused, not “free”. Retrieved from https://www.psychologytoday.com/us/blog/the-natural-unconscious/200906/the-will-is-caused-not-free

Batthyany, A. (2009). Mental causation and free will after Libet and Soon: Reclaiming conscious agency. In Batthyany, A., Elitzur, A. C. (Eds), *Irreducibly Conscious. Selected Papers on Consciousness*. Heidelberg: Universitätsverlag Winter.

Baumeister, R. F. (2009, June 30). John Bargh and some misunderstandings about free will. Retrieved from https://www.psychologytoday.com/us/blog/cultural-animal/200906/john-bargh-and-some-misunderstandings-about-free-will

Baumeister, R. F., Masicampo, E. J., & DeWall, C. N. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin*, *35*(2), 260–268.

Bode, S., He, A. H., Soon, C. S., Trampel, R., Turner, R., & Haynes, J. D. (2011). Tracking the unconscious generation of free decisions using ultra-high field fMRI. *PloS One*, *6*(6), e21612.

Bode, S., Murawski, C., Soon, C. S., Bode, P., Stahl, J., & Smith, P. L. (2014). Demystifying “free will”: The role of contextual information and evidence accumulation for predictive brain activity. *Neuroscience and Biobehavioral Reviews*, *47*, 636–645.

Breitmeyer, B. G. (1985). Problems with the psychophysics of intention. *The Behavioral and Brain Sciences*, *8*, 539–540.

Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, *6*(2), 119–126.

Dennett, D. C. (2014, January 26). Reflections on free will: A review by Daniel C. Dennett. Retrieved from https://samharris.org/reflections-on-free-will/

Edelman, G. M. (1989). *The remembered present: A biological theory of consciousness*. New York, NY: Basic Books.

Edelman, G. M. (2003). Naturalizing consciousness: A theoretical framework. *Proceedings of the National Academy of Sciences*, *100*(9), 5520–5524.

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.

Greene, J., & Cohen, J. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1451), 1775–1785.

Haggard, P. (2008). Human volition: towards a neuroscience of will. *Nature Reviews Neuroscience*, *9*(12), 934–946.

Haggard, P., & Eimer, M. (1999). On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research*, *126*(1), 128–133.

Harris, S. (2012). *Free will*. New York, NY: Free Press.

Haynes, J. D. (2011). Decoding and predicting intentions. *Annals of New York Academy of Sciences*, *1224*, 9–21.

Koch, C., & Crick, F. (2001). The zombie within. *Nature*, *411*(6840), 893.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activities (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain*, 106, 623–642.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The Behavioral and Brain Sciences*, *8*, 529–566.

Libet, B. (2002). The timing of mental events: Libet's experimental findings and their implications. *Consciousness and Cognition*, *11*, 291–299.

Libet, B. (2003). Timing of conscious experience: reply to the 2002 commentaries on Libet's findings. *Consciousness and Cognition*, *12*(3), 321–331.

Mattler, U., & Palmer, S. (2012). Time course of free-choice priming effects explained by a simple accumulator model. *Cognition*, *123*(3), 347–360.

Murawski, C., Harris, P. G., Bode, S., & Egan, G. F. (2012). Led into temptation? Rewarding brand logos bias the neural encoding of incidental economic decisions. *PLoS One*, *7*(3), e34155.

Nahmias, E., Coates, D. J., & Kvaran, T. (2007). Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy*, *31*(1), 214–242.

Roskies, A. L. (2010). How does neuroscience affect our conception of volition? *Annual Review of Neuroscience*, *33*, 109–130.

Searle, J. R. (2001). Free will as a problem in neurobiology. *Philosophy*, *76*(4), 491–514.

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.

Trevena, J. A., & Miller, J. (2002). Cortical movement preparation before and after a conscious decision to move. *Consciousness and Cognition*, *11*(2), 162–190.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science*, *19*(1), 49–54.

Wegner, D. M. (2003). The mind's best trick: how we experience conscious will. *Trends in Cognitive Sciences*, *7*(2), 65–69.