DOI 10.1007/s11055-019-00777-1
Neuroscience and Behavioral Physiology, Vol. 49, No. 5, June, 2019

The Problem of Free Will in the Context of Neuroscience Research

D. N. Razeev

UDC 612.821:159.943

Translated from Zhurnal Vysshei Nervnoi Deyatel'nosti imeni I. P. Pavlova, Vol. 67, No. 6, pp. 721–727, November–December, 2017. Original article submitted January 18, 2017. Accepted May 22, 2017.

This paper provides a critical discussion of two experiments which have been termed free will experiments conducted by scientists working in the field of cognitive psychology and neurosciences in recent decades. The author proposes two conceptual objections to the conclusion that these experiments demonstrate the absence of free will in humans.

Keywords: Libet's experiment, Soon's experiment, the "veto" principle, free will, freedom of action.

The problem of free will was and still is one of the most widely discussed not only among professional philosophers, but also among scientists. It has long been considered that the question of free will is within the domain of so-called humanities knowledge and that any attempts to answer it by the means and methods of the natural and exact sciences are doomed to failure. However, in recent decades we have seen a series of curious scientific experiments and a number of hypotheses and theories addressing this problem within the natural sciences paradigm. Three main natural science contexts in which the problem of free will is studied and discussed by contemporary scientists can be discriminated: 1) the context of physical processes linked with quantum indeterminacy; 2) the context of neurophysiological processes linked with the functioning of the human brain, and 3) the context of algorithmic processes linked with the development and construction of powerful artificial intelligence.

In this paper I focus on the second of these contexts. My task will be to lay out the two most discussed experiments carried out by scientists working in the field of cognitive psychology and neurosciences in recent decades, termed free will experiments, and to provide a critical discussion of the question of whether the data obtained from these experiments lead to a fundamentally new argument contributing to answering the classical question of free will. To simplify the discourse and to make my thoughts accessible to a wider readership, I will try to avoid any unnecessary use of the specific terminology of contemporary philosophy and will focus on the problematic presentation of the material.

Libet's Experimental Paradigm. The beginning of the current lively discussion of free will among philosophers, psychologists, and neuroscientists came from an experiment run by the American psychophysiologist Benjamin Libet at the University of California, San Francisco at the beginning of the 1980s [Libet, 1985]. The essence of the experiment was as follows. Volunteers were asked to carry out a specified body movement at a moment of their choice (for example, movement of the right index finger; I will use this version of the experiment, though in the initial version subjects were asked to move their wrist). Libet and colleagues were able to observe and fix complex neurophysiological processes occurring in the brain and muscles of each subject using noninvasive technical methods: the process of muscle movement was recorded using the electromyograph, while the preceding process in the motor areas of the brain (the readiness-to-act potential) was recorded using the EEG. With the aim of understanding at which moment the subject took the decision to carry out the corresponding body movement (raise the index finger), Libet developed the following algorithm for actions. A large clock face with a hand was placed in front of the subject; the hand moved quite slowly, making a complete rotation in about 2.5 seconds. The subject's task was to remember where the hand was positioned on the clock face at the moment at which they took the decision to carry out the corresponding body movement. Thus, Libet was able to fix not only cerebral and muscular activity, but also to place them as though on a single scale

Department of the Philosophy of Science and Technology, St. Petersburg State University, St. Petersburg, Russia; e-mail: d.razeev@spbu.ru.

with a subjective flow of the subject's experience and his or her desire to carry out the specified action on his or her own volition. The results of these experiments were somewhat unexpected. In each case, the process of taking the decision to carry out the body movement was found to be preceded by activity in the motor areas of the brain, and the subject was not aware of this. Libet established that activity in the motor cortex started 350 msec before the subject was aware of the decision to carry out the corresponding body movement. In more detail, the readiness-to-act potential started to form in the subject's brain 550 msec before the motor action itself, while the subject's desire to carry out the action was fixed only 200 msec before it started (here we note that the movement becomes irreversible 50 msec before it starts). The initial conclusion published by Libet was that each of our free acts is preceded by unconscious neural activity in the brain. This means that the decision to carry out one or another action, which is conventionally regarded as a free decision and has been regarded as the basis for the action itself, is not a voluntary (free) act, as it started as a result of another physical process occurring in the brain. The decision to carry out one or another action cannot be regarded as free, as the action starts before we take the decision, such that it has similar status to some means of fixing the initial process in the brain which with hindsight is already at the level of subjective experience. This conclusion at first sight eliminates the whole of previous metaphysical tradition, whereby consciousness is the initiator of free deeds and actions. Any of our free actions is, according to Libet's conclusions, predetermined and independent of our direct decision to perform it.

In other words, Libet's experiment provided evidence that the deeds of a human being are not the result of his or her free conscious decision, but are due to the previously determined objective processes in his or her brain occurring before its conscious phase. The conscious phase is accompanied by a characteristic type of illusion, that the initiation of these deeds came from the subject him- or herself. If this is so, then we are no more than some kind of animatronic experiencing the illusion of free will in our actions, which in reality are controlled by our brains.

Interpretation of the Results. Libet's experiment (and the interpretation of its results) became the subject of quite serious and comprehensive criticism, and this continues to date. At first, most objections were linked with the technical aspects of the experiment, particularly errors in the operation of the equipment used for measuring the readiness-to-act potential and muscle activity, as well as the problem of the consistency of the data with those obtained from subjects' reports of the time at which they took the decision. More significant objections were raised relatively recently, and these related to the experimental paradigm itself. For example, the discovery of so-called mirror neurons showed that activation of the premotor and motor areas of the cortex occurs not only when the subject carries out the specified body movement, but also when the subject observes this movement being performed by another subject. Thus, activation of the premotor and motor areas of the cortex does not always lead to the specified muscle movement so it does not provide adequate grounds to explain why the subject carried out one or another body movement. In other words, although the readiness-to-act potential recorded by the apparatus starts 550 msec before the planned body movement (for example, raising the index finger), in and of itself it cannot be regarded as the cause of the corresponding body movement. This requires some sort of additional mechanism, which some authors call the "conscious proximate decision" (see, for example, Mele, 2009; 2013], which occurs (as indicated by the reports of the participants in Libet's experiments) 200 msec before the corresponding body movement.

It should be noted that Libet himself, aware of the radical nature of the consequences flowing from his experiment, decided to correct his theoretical scientific conclusions, leaving a special place for free will, as a result of more detailed observation of the course of his experiment and minor modifications. Multiple repetition of the experiment showed that subjects did not always complete the action for which the readiness potential had started to form in the brain and which was predicted by the experimenter. Some subjects took the decision not to perform the action at the very last moment. Libet's interpretation was that this means that the subject whose brain had started to form the readiness potential prior to the planned action had, after it came to consciousness (i.e., 200 msec before it started) but before the moment at which the action became irreversible (i.e., 50 msec before it started), a period of around 100–150 msec in which to cancel the decision, which had already been taken in the brain, to carry out the action. Libet termed this mechanism the "veto principle" and suggested that this is what corresponded to free will. In fact, if this curious conclusion is considered in detail, Libet proposed that free will should be identified with negative freedom, i.e., the subject's right to consciously apply a veto to those decisions taken by the brain. This conclusion by the American scientist can be regarded as a certain compromise, proposed to naturalize one of the main phenomena conventionally associated with consciousness, i.e., free will. Thus, it seems that we are some kind of semi-animatronic with a reverse mechanism. On the one hand, all our deeds are controlled by our brain and the processes occurring within iton the other hand, some of those actions which are planned by the brain before reaching consciousness and for which specific trigger mechanisms have already been launched can be deflected by the subject at the stage of consciousness, all the way to the moment at which they become irreversible; in other words, we are to some extent free not to perform them.

The Soon et al. Experiment. Two decades after Libet's original experiment and the scientific/philosophical discussions which accompanied it at all significant stages of its

reproduction, another series of interesting experiments was conducted at the beginning of the 21st century in Germany, which led to discussion of free will at a new level [Soon et al., 2008]. In Soon and Haynes' laboratory, volunteer subjects were asked to make a free choice of an elementary operation – which button on a panel to press: the one on the right or the one on the left. At the same time, the subject saw a series of changing letters of the alphabet on a monitor screen. The subjects' task was to remember which letter was on the monitor screen at the moment at which they took the decision as to which button to press (left or right). Neuron activity in the subjects' brains was recorded by fMRI. The essence of the experiment, in the organizers' words, consisted of decoding the intent to select the right or left button as it appeared in the subjects' brains. The neuroscientists obtaining the first results on neuron activity in the experimental participants asked programmers to write a special computer program to predict, using nothing more than the fMRI data, which of the buttons on the panel the subject would choose before he or she declared it. During the experiment it was noted that the computer program, using patterns of neuron activity, could reliably predict a subject's "free" decision before the subject announced it in more than 60% of cases. The results of this experiment evidently confirmed Libet's view that neuron activity in the motor areas of the brain and not the consciously taken decision was the initiator of one or the other action by the subject. However, new technical abilities not available to Libet a quarter of a century previously allowed a new paradigm to be created to confirm its conclusions. In contrast to Libet's experiment, where the discussion centered on milliseconds, the results of the Berlin group's experiments were simply staggering. The new experimental paradigm allowed a subject's future choice to be predicted 6–10 sec before the subject made the choice! Despite the fact that the program made correct predictions in only 60% of cases, the result was nonetheless greater than by simple chance and it could be influenced, as suggested by the scientists themselves, by a number of attendant factors such as errors in measuring neuron activity, the noise level, etc. Regardless, fixation of this long delay between the processes in the brain initiating the action and the "free" decision to perform it meant that the large number of opponents could not regard Libet's conclusions that there is no positive freedom of volition in humans as incorrect on the grounds of the inaccuracy of and errors in the apparatus used in running experiments of this type.

Conceptual Objections. Some scientists have come to the radical conclusion that free will is illusory, on the basis of results from one or both of the experiments described above [Wegner, 2002; Singer, 2004; Haggard, 2008; Frith, 2009; Coyne, 2012]. In addition, serious progress in the techniques used to measure neuron activity when people carry out voluntary movements which have made these experiments possible have not removed a number of fundamentally important questions from the agenda: does this kind of experiment really hit the target? Can we take the view that experiments on voluntary movements of some part of a person's body concern the problem of free will? Is it correct to regard raising or not raising a right index finger or the choice between which of two buttons to press in the conditions of a controlled experiment as free will?

I would like to put forward two conceptual objections to the conclusion that there is no free will in humans made on the basis of the experiments described above. I regard these objections as conceptual, as they have allowed and continue to allow scientists to change their understanding while running and interpreting experiments. I call the first of these objections the weak conceptual objection, and the second the strong conceptual objection.

The weak conceptual objection is as follows: the experiments laid out above are experiments with freedom of action, but not freedom of will. One of the first to distinguish between freedom of will and freedom of action was the American philosopher Robert Kane, an authoritative expert in studies of the problem of free will [Kane, 1996]. I will not reproduce his argument in detail, as I understand the distinction somewhat differently. From my point of view, freedom of will and freedom of action are fundamentally different types of freedom. Freedom of action is linked with implementation of a first-order intention, while freedom of will is related to implementation of second-order (reflective) intention. I use the term "first-order intention" to refer to direct intentions (for example, "I want to raise my left arm" or "Tomorrow I will get up at eight o'clock," or even "When I finish school I'm going to go to university;" these are united by the fact that the content of these intentions is the specific state of affairs to which they are targeted). Second-order intentions are indirect intentions, whose content consists of our first-order intentions (for example, "I want to quit smoking," "I want people to love me," "I'm never going to tell lies," "I refuse to turn a blind eye to injustice"). The American philosopher Harry Frankfurt introduced a similar distinction, namely that between first-order desires and second-order desires [Frankfurt, 2003]. In addition, from my point of view desire is a subtype of intention, so I prefer to build the conceptual objection on differences between first- and second-order intentions. Implementation of first- and second-order intentions requires different types of freedom. If freedom of action is sufficient for implementation of first-order intentions, freedom of will is needed for implementation of second-order intentions. The fundamental difference is that first-order intentions are discrete, they terminate at the moment at which they are fulfilled, and the desired state of affairs is achieved. In contrast, second-order intentions are continuous, i.e., they retain their action in the individual's mind. At the phenomenological level (the self-observation level), the difference between freedom of will and freedom of action can be identified. For example, when I smoked I repeatedly identified the difference between my freedom to act and my freedom to want. Although I was

free to implement first-order intentions (I was free to choose what to smoke – cigarettes, cigarillos, or even cigars), I long lacked the freedom to implement the second-order intention (the wish to quit smoking). Furthermore, despite the fact that I quit smoking many years ago, implementation of this second-order intention (to quit smoking) is not complete and lasts to this day, requiring continual maintenance (freedom of will). Identifying the difference between freedom of will and freedom of action at the phenomenological level, I propose that their correlates may be different mechanisms of neuron activity in the brain. Thus, the intention of freedom of action and the intention of freedom of will may also require different measurement techniques, i.e., they may be found in different experimental paradigms. Thus, to conclude that we have no freedom of will, scientists need to conceive and create an experimental paradigm in which measurements are made not only of the first-, but also the second-order intention. Progress in techniques of measuring the intentions of freedom of action do not automatically mean progress in studying the intentions of freedom of will. It is entirely possible that measurements of readiness potentials are not at all suitable for solving the challenge of free will.

If we accept that these experiments were experiments not on freedom of will but on freedom of action, would it be correct to conclude that they provide evidence that humans lack freedom of action? At this point I will move on to the next objection, which I term the strong conceptual objection.

The strong conceptual objection is as follows: in these experiments, the scientists measured and interpreted subjects' experiences of the motivational experiences type rather than the intentional type. The fact is that in conducting these experiments, the scientists did not provide any clarification as to the type of mental experiences they were linking with performance of voluntary movements. Considering freedom of action, the scientists put desire, purpose, impulse, decision, intention, and so on in a single row, as though these concepts meant essentially the same thing. However, can we regard, for example, a sudden impulse to move the right index finger as the same as the long-nurtured desire to undergo sex reassignment surgery? From my point of view, we are dealing with fundamentally different types of experiences: in the former case, the discussion relates to a motivational experience (the impulse to do something), while in the second it relates to intentional experience (the intent to do something). Motivation and intention cannot be regarded as interchangeable concepts in describing our freedom of actions, as they have different conceptual contents. I will clarify my meaning using an elementary example. Imagine that one of my best friends invites me to a restaurant and another invites me to the cinema (at the same time). Two impulses can arise in my mind – to go to the restaurant with one friend and to go to the cinema with the other. However, if I am in my right mind, two intentions could hardly occur in my head at the same time, i.e., to go to the restaurant with one friend and at the same time to go to the cinema with the other. In other words, at the conceptual level, the concept of "impulse" allows the coexistence of two different impulses in the subject's mind at the same time, though the concept of "intention" does not. If motives and intentions differ conceptually, they may not be identified in experiments and their interpretation. From my point of view, at the phenomenological level, we can identify the difference between intentional freedom of action and motivational freedom of action. I believe that the neuroscience and cognitive psychology experiments described above addressed motivational freedom of action. It is entirely reasonable to suppose that the mechanism of neuron activity underlying motivational freedom of action is significantly different from the mechanism correlating with intentional freedom of action. Even if some neuroscientists believe that these experiments demonstrated that voluntary movements involve activation of the motor areas of the brain preceding the "decision" to make these movements, the conclusion that there is no freedom of action is premature, as these studies addressed only one type of freedom of action, i.e., that linked with motives but not that linked with intentions.

From my point of view, scientists need to provide a clearer definition of their concept of a free action before making far-reaching conclusions regarding determinism and the lack of freedom of our actions. If there is no clear understanding of what a scientist wants to measure in some experiment, then no matter how technically perfect the experiment is, it will to some extent be random and its result will remain a cryptic set of data, zeros, ones, and other symbols.

Conclusions. Thus, scientists' experiments have not demonstrated the lack of free will in humans, as the conceptual apparatus used by scientists in devising these experiments contains a double substitution of concepts (freedom of will is replaced by freedom of action, while freedom of action is restricted to motivational freedom of action).

In addition, this conclusion does not mean that free will arises from some non-material source and is fundamentally incapable of technical measurement in terms of neuron activity. I am convinced that if free will exists, then it is as a physical phenomenon, like other physical phenomena in the world. However, I take the view that no experiments with freedom of will have been conducted in the recent history of the neurosciences and cognitive psychology. Furthermore, the possibility of devising such an experiment requires serious consideration.

I would like to finish this paper with a single important remark, which may somewhat disappoint the reader. However, I am unable to do so. Do my conceptual objections mean that free will exists? No, this is not the case. All I wish to show is that scientists' experiments do not demonstrate the *absence* of free will. Thus, my belief is that there is only a possibility that free will exists. However, that does not mean that free will has to exist. My point of view on this point can be characterized as guarded optimism. I believe that the question of free will is not only a philosophical problem, but also an empirical problem, and that it is appro-

priate for it to be solved in the framework of the natural sciences. This requires a new empirical paradigm, i.e., consideration of experiments addressing the question of free will in a more adequate way than neuroscientists and cognitive psychologists have achieved to date. This paradigm could be created on the pathway to a productive collaboration of scientists and humanists.

REFERENCES

Coyne, J. A., "Why you don't really have free will," *USA Today*, Jan. 2, 2012,www.usatoday.com/news/opinion/forum/story/2012-01-01/free-will-science-religion/52317624/1.

Frankfurt, H., "Freedom of the will and the concept of a person," in: *Free Will*, G. Watson (ed.), Oxford Univ. Press, Oxford (2003).

Frith, C., "Free will top-down control in the brain," in: *Downward Causation and the Neurobiology of Free Will*, N. Murphy, G. F. R. Ellis, and T. O'Connor (eds.), Springer Verlag (2009), pp. 199–209.

Haggard, P., "Human volition: towards a neuroscience of will," *Nat. Rev. Neurosci.*, **9**, 934–946 (2008).

Kane, R., *The Significance of Free Will*, Oxford Univ. Press, New York (1996).

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

Mele, A., *Effective Intentions*, Oxford Univ. Press, Oxford (2009).

Mele, A., "Free Will and Neuroscience," *Philosophic Exchange*, **43**, No. 1, Art. 3 (2013).

Singer, W., "Verschaltungen legen uns fest: Wir sollten aufhören, von Freiheit zu sprechen," in: *Hirnforschung und Willensfreiheit. Zur Deutung der neuesten Experimente*, C. Geyer (ed.), Suhrkamp Verlag, Frankfurt am Main (2004).

Soon, C., Brass, M., Heinze, H.-J., Haynes, J.-D., "Unconscious determinants of free decisions in the human brain," *Nat. Neurosci.*, **11**, No. 5, 543–5 (2008).

Wegner, D. M., *The Illusion of Conscious Will*, MIT Press, Cambridge (2001).