Critical Reasoning 27 – Experimental Philosophy

for Ammu

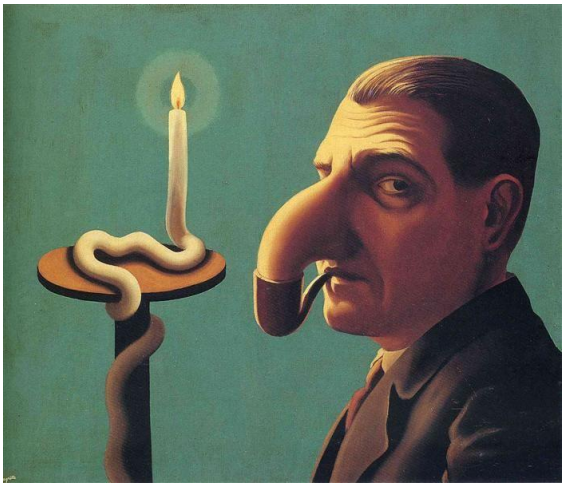
Joshua Knobe: Experimental Philosopher and Professor of Cognitive Science and Philosophy at Yale University

Experimental philosophy refers to an emerging field of interdisciplinary research that makes use of empirical data, gathered through surveys or experimental techniques, in order to inform philosophical enquiry. Since Plato philosophers have relied on *a priori* justification and rational argument, often by means of analogy. However since the Scientific Revolution and Enlightenment, and onward towards the 21st Century, ordinary intuitions about such matters as human nature, ethics, causality and even space and time have sometimes proven to be inconsistent with the best empirical evidence offered by the sciences. Philosophers of all stripes have therefore been at pains to take cognoscence of developments within the sciences to the extent that they overlap with their fields of interest. Experimental philosophers however go further. If an intuition is equivocal or if there are lacunae in the evidence base that might inform some philosophical theory, they seek out decisive evidence by empirical means, including experimentation, if it is feasible.

There are two articles that inform the discussion that follows: the first and more recent by Joshua Knobe (2011) and the second by Benjamin Libet (1985). Both may be downloaded here and here for free. Note that South African copyright law allows for the reproduction of individual journal articles for educational purposes.

Armchair Philosophy

We are all familiar with one of the stereotypes of philosophy: a man (and it is always a man) seated in a comfortable armchair, perhaps with a pipe, deep in thought about some abstract or theoretical matter, paying little head to the world about him. At its best this sort of philosophy involves a careful analysis or synthesis of existent scholarship but at the other extreme may simply be a frivolous or superficial rehashing of old information or theories. The latter was arguably satirised by the surrealist artist Rene Magritte in his 1936 *The Philosopher's lamp*, at right. Note the way the snaking candle illuminates the philosopher smoking his own nose.



© philosophy.org.za

Of course very few historical philosophers have been unconcerned with empirical evidence, although even Aristotle (see Critical Reasoning 11) was inconsistent, at times failing to check the most basic facts, at others providing the most detailed biological observations.

Thought Experiments

As we saw in Classic Text 22, both philosophers and theoretical physicists use thought experiments which are "science fiction" like scenarios imagined in order to explore the logical relations, implications and compatibility between scientific theories and data as well as among other theories. Thought experiments therefore have the dual advantage of being able to conceptually "try out" ideas before empirical data become available, as well as broadening the domain of those theories in which they feature. In this sense, thought experiments provide a bridge between speculative armchair type philosophy or physics and testable hypotheses that can be incorporated into such theories.

Scientific and Philosophical Instruments

Knobe (2011) observes that within a typical scientific research programme, scientists work with certain instruments, such as telescopes and microscopes. These two in particular broaden the scientist's scope of observation to from light years on the one hand down to 0.01 nm depending on the microscope used on the other hand. Of course, scientists do care about the instruments they use, if only because they are so costly and delicate. They also spend a good deal of time servicing and calibrating them. Ultimately though, they are just tools to gather information about some independently existing reality. However when information coming from the instruments is puzzling, implausible or confusing it is helpful to turn away from the reality one is trying to represent and look instead in detail at the instruments themselves. Was the puzzling output from the radio telescope not caused by its detecting a signal from a quartz wrist watch worn by one of the observers? Or was the streak seen on a slide not an artifact of a blunt microtome blade used in its preparation?

At any rate, philosophers do not make much use of such instruments. Instead they mostly employ the human mind as an instrument that represents a mediated internal reality. Like scientists, most philosophers, until very recently, have not been overly concerned about the detailed workings of human minds, relying on them to get a handle on independent reality. Sometimes however this approach fails or produces equivocal answers to the same question. In such situations it can be helpful to explore the mind itself and look scientifically at the source of our own and other's philosophical intuitions. This is where experimental philosophy provides the key. If researchers can gain an understanding of the psychology motivating philosophical intuitions then we will be in a better position to judge which intuitions are trustworthy and which are unreliable or misleading. According to Knobe:

> This work, we hope, will give us a better understanding of people's beliefs about the great philosophical issues. How is it that individuals come to believe in free will? Do they see their own moral claims as objective truths? The findings could ultimately have practical implications in jurisprudence, ethics and other fields. (p. 58)

© philosophy.org.za

Free Will: One Experimental Approach

Imagine that you are a witness to a murder. It may seem immediately obvious that the murderer is morally (and criminally) responsible for his actions and that he is deserving of punishment. However, as you later mull the matter over philosophically, it may occur to you that the murderer's actions were cause by certain mental states, which presumably were caused by prior states and events. Perhaps the actual killing was just the final outcome of a long chain of events that could be traced back to his genes and the environment that shaped him. If so, could he really be morally responsible for the things he has done? Philosophers of different persuasions give different answers based on their conception of free will. Perhaps our capacity for abstract theoretical reflection pulls us one way, while our emotional response to the horror of the deed pulls us in the other direction. Knobe & Nichols (2008) describe their experiment designed to test these intuitions.

At the start of the experiment all participants were asked about a fictitious universe called 'Universe A' in which everything anyone did was completely determined by a causal chain of events stretching back into the past. Participants were then randomly divided into two groups. Those in one group were then asked a question designed to trigger abstract theoretical reflection:

> In Universe A, is it possible for people to be fully morally responsible for their actions?

Those in the other group were given a highly concrete, even lurid description designed to elicit a more emotional response:

> In Universe A, a man named Bill is attracted to his secretary and decides that the only way to be with her is to kill his wife and three children. He knows that it is impossible to escape from his house in the event of a fire. Before leaving on a business trip, he sets up a device that burns down the house and kills his family. Is Bill fully morally responsible for killing his wife and children?

According to Knobe (2011) those who were asked the abstract theoretical question tended to say that no one can be morally responsible in a deterministic universe, while those who received the more concrete description tended to take the opposite view, saying that Bill was indeed responsible for his actions. Although the original experiment was a small one, conducted with a few dozen American undergraduates, a number of studies over the following years explored the phenomenon more rigorously. One such experiment was conducted with more than 1000 participants; another explored intuitions about the abstract case across different cultures in India, Hong Kong, Columbia and the USA. Each time the original findings were confirmed, yet questions remained as to why this occurred. "Did the effect actually reflect a difference between abstract and concrete thinking?" (p. 58)

An elegant test devised by experimental philosopher Christine Weigel provided further insight. Weigel's participants we asked to imagine hearing a philosophy lecture about the problem of free will which began in a general way and concluded with the example above *i.e.* a man in a deterministic universe who kills his wife and children. Weigel however introduced a subtle difference. Some of the participants were asked to imagine the lecture occurring "in a few years", while others were asked to imagine the lecture occurring "in a few days". While this difference apparently has little to do with the philosophy of free will, it has been shown that it does lead people

to employ different cognitive processes. Imagining events in the distant future is associated with more abstract, reflective, theoretical and high level cognitive processes; whereas imagining events in the immediate future triggers more concrete intuitions. This is precisely what Weigel found: those asked to imagine the lecture occurring in the distant future were more likely to say that individuals in a deterministic universe were not morally responsible for their actions. (p. 58)

Although these experiments were only a beginning, it does seem that people's ambivalence towards the problem of free will derives from a conflict between more abstract, theoretical judgements and their more concrete, emotional responses. Knobe however mentions a rival hypothesis by the experimental philosopher Eddy Nahmias that does not involve any conflict between reason and emotion. At the time, Knobe (2011) claimed "that the evidence available now from these experiments is not sufficient to resolve all the major questions… Although a great deal still remains to be done, we now have the beginnings of an experimental research program on the psychological roots of people's understanding of free will". (p. 59) What is refreshing about the experimental approach is that philosophers no longer have to retreat to their armchairs – they have a chance to seek out the evidence.

Is Morality Relative?

Another perpetual question which is easy to argue but difficult to demonstrate is whether or not morality is relative to culture. We have consistently maintained that moral relativism is inconsistent and self-defeating; however this is not what most "politically correct" activists proclaim. According to the latter, there is no single moral truth because morality is always fundamentally relative. Conservative authorities, on the other hand, insist on the existence of objective moral truths. Pope Benedict XVI, for example, declared that relativism leads "to moral or intellectual confusion, to a lowering of standards, to a loss of self-respect, and even to despair". (p. 59)

In order to investigate the controversy psychologist Edward Cokely and philosopher Adam Feltz conducted an experiment in which they gave participants a story about people who hold opposite views on a moral question. Participants were then asked whether one disputant had to be wrong (the anti-relativist response) or whether there might be no single correct position (the relativist response). Participants were also given a standard assessment of "openness to experience", one of the "big five" personality traits. The results showed that the higher a participant scored in openness to experience, the more likely that they were to endorse the relativist answer. The results suggest the following hypothesis: people who are attracted to moral relativism may feel that they are open to other ways of belief and are willing to engage in them imaginatively. (p. 59)

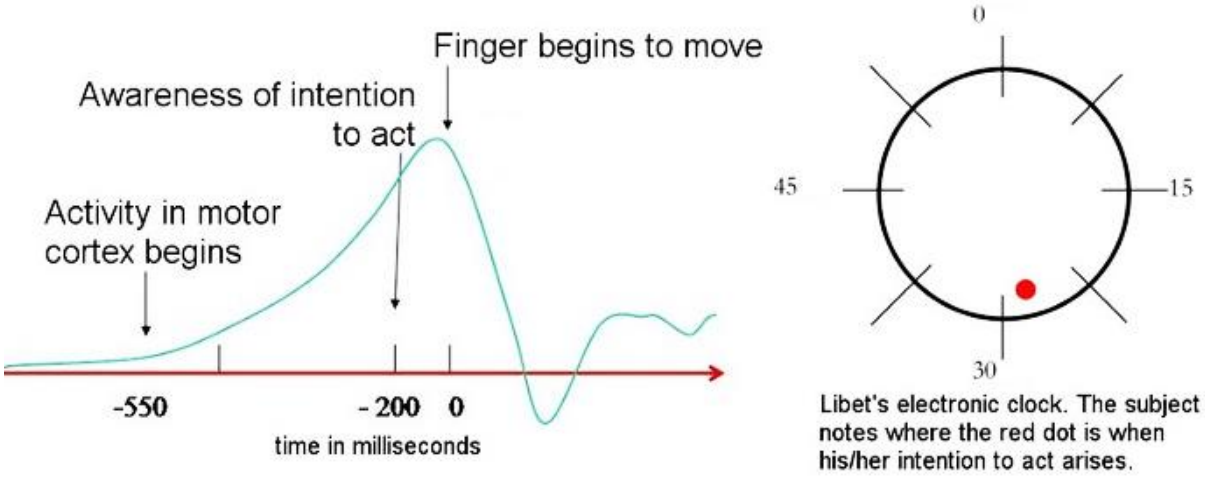
Psychologists Geoffrey Goodwin John Darley devised an indirect test of this hypothesis by asking participants to solve a logical puzzle involving configuring blocks in a certain way. According to Knobe, although the puzzle seemed straightforward, it could only be solved by regarding the problem from multiple perspectives. The research hypothesis then was about people's ability to solve such puzzles and their intuitions about relativism. It turns out that both were significantly correlated. Those who solved the problem were more likely to provide relativist answers. So various studies appear to converge on the same conclusion: that people feel drawn towards relativism to the extent that they are open to other perspectives. The results do not solve the problem of moral relativity; however they do provide a window onto psychological roots of the controversy. (p. 59)

© philosophy.org.za

Free Will: Libet's Experiment

A classic experiment concerning the neuroscience of free will conducted by Benjamin Libet and colleagues in 1983 (Libet 1985) has been both celebrated and criticised by neurologists, psychologists and philosophers from diverse backgrounds. In the experiment participants were asked to flex their wrist or the fingers of their right hand at any time the wished while researchers simultaneously recorded their brain electrical activity via an electrode on the scalp just over the motor/premotor cortical area that controls the hand. At the same time another electrode above participants' muscles responsible for the flexation of the wrist and fingers registered their contraction. Two types of timing of events were recorded: the objective time registered via the electrodes and a subjective "clock time" as self-reported by the participants. In the case of the latter, each participant looked at a spot which revolved on a clock face at a quick but steady rate before them. They reported more or less where the spot had been the moment they felt the desire to flick their wrist or fingers. Participants were not pressured or constrained to act within a given timeframe; therefore their timing was spontaneous and fully endogenous.

Almost two decades before Libet's experiment Hans Helmut Kornhuber and Lüder Deecke (1965) reported a negative spike in recorded neural activity of the motor cortex and supplementary motor area that preceded voluntary movement by 1 - 2 seconds among humans. This **readiness potential** (**RP**) as it is now known can clearly be seen on an electroencephalograph (EEG) readout, whether on graph paper or an electronic screen. Libet and his team wanted to know whether the RP corresponded to the felt intention to produce a movement. If that were the case, they reasoned, the RP should occur at the same "clock time" that the subject reported feeling the urge flick their wrist or fingers. Instead the team found that that there was unconscious brain activity including the RP, as recorded by the scalp electrode, approximately 500 milliseconds (ms) before subjects became aware of their conscious decision to move. In other words, activity in the subjects' motor cortices associated with the movement began *before* they became aware of their intention to act. See schematic diagram below.



The vertical axis is measured in tens of millivolts (not shown and not to scale) where the upward direction is negative. The horizontal axis is measured in milliseconds. Activity in the motor cortex begins at least 550ms before the fingers begin to move; however awareness of the intention to act as timed by noting the position of the dot on the clock face is only registered some 200ms before the act at time zero.

© philosophy.org.za

Libet's counterintuitive results suggest that the intention to move begins unconsciously, well before the conscious awareness of the subject to do so, and that the subject's belief that it occurred at their behest was only a retrospective perspective on the event. Libet's primary findings do not imply that free will is impossible, only that the initiation of spontaneous voluntary acts in the brain begins unconsciously and only later reaches conscious awareness. Another important finding, not shown above, was that in the interval 100 - 200ms before the planed time to act, subjects were able to censure or veto their intention and thus supress the action. Libet (1985 p. 539) concludes:

> The concept of conscious veto or blockade of the motor performance of specific intentions to act is in general accord with certain religious and humanistic views of ethical behavior and individual responsibility. "Self-control" of the acting out of one's intentions is commonly advocated; in the present terms this would operate by conscious selection or control of whether the unconsciously initiated final volitional process will be implemented in action. Many ethical strictures, such as most of the Ten Commandments, are injunctions not to act in certain ways. On the other hand, if the final intention to act arises unconsciously, the mere appearance of an intention could not consciously be prevented, even though its consummation in a motor act could be controlled consciously. It would not be surprising, therefore, if religious and philosophical systems were to create insurmountable moral and psychological difficulties when they castigate individuals for simply having a mental intention or impulse to do something unacceptable, even when this is not acted out (e.g., Kaufinann 1961).

Daniel Dennett (1991) called Libet's findings into questing, claiming *inter* alia that subjects would have had to shift their attention from their internal clock to the external one and that that this would have introduced temporal mismatches between the felt experience of will and the perceived position of the dot on the clock face. (Dennett 1991, Gregson, 2011) Subsequent investigations have shown that the exact timing does indeed depend on attention (Haggard & Eimer 1999; Trevena & Miller 2002); however Libet's main finding have stood up to further scrutiny. (Soon et al. 2008; Haggard 2005 and Banks & Pockett 2007)

Should We Burn the Armchair?

Knobe (2011) concludes his article with this very question. Although he is photographed above wearing a T-shirt with the image of a burning armchair, this does not reflect his more serious attitude. It may be that one day most of our empirical questions will be resolved by experimentation, and that we will eventually arrive at an accurate understanding of the cognitive processes underlying most people's philosophical intuitions. However, even then we will not have exhausted the most important philosophical issues, such as whether our intuitions are actually right or wrong, or what it means for us to be free agents. "Sooner or later someone is going to have to get back into that armchair and reflect hard on the philosophical issues themselves." Philosophical refection and experimental philosophy are not alternatives; rather the latter should add another tool to the toolbox of philosophical inquiry, including dispassionate observation, logic, analogy and thought experiments. (p. 59)

Daniel Dennett, whom we met in Classic Text 14, is one philosopher who has consistently payed attention to empirical findings in developing his thinking. In Dennett 2003 Ch. 8 under the heading

© philosophy.org.za

Whenever the Spirit Moves You (p. 227 ff) Dennett discusses the philosophical assumptions and interpretations of Libet's findings. These include the nature of the observing "self" and judgements about the observed simultaneity of events. For Dennett, "What Libet discovered was not that consciousness lags ominously behind conscious decision, but that conscious decision-making takes time. If you have to make a series of conscious decisions, you'd better budget half a second, roughly, for each one, …" (p. 239)

Task

1. Is there a limit to what can be studied by experimental philosophy?

2. What are the philosophical implications of Libet's experiment for the understanding of the problem of free will?

Feedback

1. If, as we have argued in Classic Text 28, there are limits as to what can legitimately be studied by philosophy then there is a much narrower limit as to what can legitimately studied by experimental philosophy. Experimental philosophy relies on gathering empirical data, whether in the physical or psychological sciences. While surveys may be made of the intuitions of ordinary people, the justification for such intuitions, if any, are much harder to capture. Concepts involving values or judgements, whether aesthetic or moral, or related to meaning, whether semantic or existential, do not lend themselves to empirical study.

2. Your answer will depend very much upon the assumptions you make about free will and determinism. Libet (1985) claimed that his definition of voluntary action "is not committed to or dependent upon any specific philosophical view of the mind-brain relationship". (p. 530) Nor did he deny that we do act voluntarily, some of the time, only that there is an early unconscious component to such acts. Neither did he claim, as some have glibly said, that we "don't have free will, but we do have free won't". Consulting Dennett (2003 Ch. 8) as well as revising Classic Text 14 should be helpful. Consider our intuitions regarding the unconscious *vs.* conscious dimension compared with involuntary *vs.* voluntary action. What Libet's experiment seems to confirm is that for some simple, voluntary motor actions at least, the intention to act begins unconsciously before the conscious awareness to do so. This seems to fit with our everyday experience "discovering" what we intend when we catch ourselves already in the act. It is not uncommon for us to confabulate conscious reasons for actions that began before we could have even been aware them. This does not mean that we are not free, only that free choice seems to be smeared out across unconsciousness and consciousness with no objective boundary between the two.

   Consider the curious, but familiar, experience of getting up to attend to some matter in another room, but that when we got there, we were stumped for why we came. Clearly, our unconscious intention was formed freely and we must have walked purposefully from one room to another to take care of what we intended; however what should have been the

© philosophy.org.za

conscious component of our action either failed to enter consciousness, or was very quickly forgotten.

References

BANKS, W. & POCKETT, S. (2007) Benjamin Libet's work on the neuroscience of free will. **In** M. Velmans and S. Schneider (Eds.) *The Blackwell Companion to Consciousness*. Blackwell

DENNETT, D. (1991) *Consciousness Explained*. Ch. 6. Penguin

DENNETT, D. (2003) *Freedom Evolves*. Viking Books, Penguin edition (2004)

GREGSON, R. (2011) Nothing is instantaneous, even in sensation. *Behavioral and Brain Sciences* **15** (2): 210 – 211

HAGGARD, P. (2005) Conscious intention and motor cognition. *Trends in Cognitive Sciences* **9**(6): 290-295

HAGGARD, P. & EIMER, M. (1999) On the relation between brain potentials and the awareness of voluntary movements. *Experimental Brain Research* **126**(1): 128-133

KAUFMANN, W. (1961) *The faith of a heretic*. Doubleday

KNOBE, J. (2011) Thought Experiments. Scientific American **305**(5) p. 56-59

KNOBE, J. & NICHOLS, S. Eds. (2008) Experimental Philosophy. Oxford University Press

KORNHUBER, H. & DEECKE, L (1965) Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Archiv für die Gesamte Physiologie des Menschen und der Tiere* **284**: 1-17; Englisch translation: (2016) Brain potential changes in voluntary and passive movements in humans: readiness potential and reafferent potentials. *Pflügers Archiv: European Journal of Physiology* **468**(7): 1115-1124. doi:10.1007/s00424-016-1852-3. PMID 27392465.

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

SOON, C. S. *et al*. (2008) Unconscious determinants of free decisions in the human brain. *Nature Neuroscience* **11**(5): 543-545

TREVENA, J. & MILLER, J. (2002) Cortical Movement Preparation before and after a Conscious Decision to Move. *Consciousness and Cognition*. **11**(2): 162-90, discussion p. 314-25

© philosophy.org.za