

Getting Carried Away With Libet **Confabulation and Conscious Will**

Duncan Crowe

1. Introduction

Benjamin Libet's experiments on the relationship between the conscious experience of volition and the onset of neural activity in the pre-motor cortex has captured the imagination of many people. One interpretation of Libet's experiment has it that Libet has shown 'free will does not exist' or otherwise that Libet has show will to be an epiphenomenon of the process of acting. I shall reject this interpretation of the experiment. In its place I shall offer an interpretation of Libet's results which draws on modern compatibilist theories of will and attempt to motivates an argument that Libet's experiments merely add to a mounting body of evidence which points to the confabulatory nature of consciousness and the limitations of our capacity for introspection.

2. Libet's Experiment and the Naïve Interpretation

Benjamin Libet sought a way to experimentally measure the relationship between 'free will' and neural activity in the brain areas associated with motion. Libet's experiments¹ consisted of subjects flexing their wrists and reporting the time at which they consciously decided to make the movement. To obtain a precise measure of the onset of wrist movement Libet used an electromyogram to monitor the contraction of the muscle cells. To obtain a measure of conscious awareness without using motor-based responses Libet had subjects look at a modified clock made from oscilloscope bulbs which cycled once every 2.4 seconds. In pre-experiment testing subjects were asked to recall and report the onset of a visual stimulus. They were found to do so with a very consistent delay of only around 50ms from the real time. It seems a reasonable working hypothesis that the perception of an internal process will be subject to even less delay than the conscious detection of a visual stimulus. Libet also set up an electroencephalogram (EEG) to monitor electrical activity in the pre-motor brain regions associated with planning an action. This gives a measures of 1) the onset of motor activity, 2) the onset of the planning of motor activity and 3) the awareness of the decision to trigger the onset of the planning of motor activity. Common sense would suggest that 3 precedes 2 which precedes 1. Common sense is wrong.

Libet found that while subjects reported the conscious awareness of the act of willing as occurring about 200ms before the onset of motor activity, the activity in the pre-motor cortex began on average 350ms before the action. This means that the neural precursors to the action taking place occur 150ms (on average) before the agent is conscious of their own decision to act. Even allowing for the systematic error associated with estimating the occurrence of conscious sensations this is at the very least a delay of 100ms between beginning to act and consciously being aware of deciding to act. Unambiguously, conscious awareness of deciding to act occurs after the action (in the broadest sense) has already begun.

¹ See Libet (1983), (2002).

The naïve interpretation of these results is to say that they show free will not to exist or, if it exists, that it is epiphenomenal to the process of acting, which is to say that it is caused by and does not cause actions to take place – regardless of whether it plays a subsequent causal role later². The argument is something to the effect that; ‘free will is a conscious process by definition, and consciousness (assuming we rule out backwards causation) as evidenced by Libet’s experiment plays no role in the production of action, therefore free will does not exist. Libet does not offer quite the same degree of pessimism himself, but does affirm the contention that the will is not the originator of action, simply adding the caveat that he sees a role for conscious will in adverting a course of action already initiated. The idea is quite simple; just because the conscious experience of willing is subsequent and not prior to the neuronal activity which plans the action by 100 to 150ms, it still appears 200ms before motor activity actually begins, which gives it a window to inhibit or prevent action, even if it does not initiate it.

3. Rejecting the Naïve Interpretation: Modern Compatibilism

The problems with the Naïve interpretation originate from two key assumptions which it makes. The first assumption is that the mechanisms referred to by talk of free will are identical with the conscious awareness of free will and that there are no forms of free will other than those which are consciously experienced (i.e. there are no unconscious mechanisms of willing, worthy of the name ‘free will’). We will call this the presumption of identity. The second assumption is that conscious experience in general is as it appears to be upon introspection. Mental events which seem upon introspection to be simultaneous are simultaneous; content which seems to be present upon introspection is present, even if it cannot be clearly articulated. These assumptions are closely related and overlap. The first is found in the naïve interpretation when it claims that by Libet having shown the conscious experience of free will to follow the initiation of action, he has showed free will to follow it also. The second is found in the idea that the conscious experience of willing must accompany any action apt to be called ‘consciously willed’ simultaneously, and that a delay of the awareness of willing entering consciousness indicates that an act was taken external to the agent’s central executive. As these presumptions overlap, by attacking the former we can use an interpretation of Libet’s experiment to make an attack upon the latter which is a much more central issue in the philosophy of mind.

The presumption of identity sets itself as far as I can see dead against sophisticated modern takes on free will from the compatibilist camp. Modern compatibilism sets itself out to argue against the very kind of free will which Libet’s experiment is taken to have disproved; namely the potent interjection of consciousness into the otherwise mechanistic decision making process. Instead, modern compatibilists³ have it that, it is from the actions of this mechanistic decision making process, honed through evolutionary processes favouring a flexible and intelligent responsiveness to the world, that the free will we have emerges. As such, while the planning of action (a decision making process) might happen to precede the conscious awareness of a

² Just to mention in passing that epiphenomenalism about volition as regards acting does not imply even partial epiphenomenalism about mental states.

³ I have in mind here the likes of Wegner, Ainslie and particularly Dennett.

decision to act, that is an interesting fact about a delay of conscious awareness; it is not a fact about a delay of decision making.

A question which is often asked here is whether or not compatibilist 'free will' is, in fact, free will. For the sake of concision my response to this shall follow Dennett's. If you insist on 'free will' as being *causa sui* and originating outside of the normal mechanics of decision making then, yes, it would appear that neuroscience cannot support your conception of free will and neither can a deterministic physicalist picture of the world. On the other hand, if you are tracing the concept of free will according to certain functions that 'free will' is supposed to perform, then (so the claim goes) this is entirely apt to be identified with the evolved mechanisms for 'eviting' undesirable outcomes in the external world and promoting desired goals. This, in effect, means there is a thing in place which does everything free will was purported to do – 'the only free will worth wanting' to use Dennett's phrase.

Suppose we consider this theory in light of the perennial issue of moral responsibility. Free will is often held as a necessary condition for moral responsibility. Incompatibilists either drop this condition or hold that, as the *causa sui* brand of free will does not in fact exist, talk of moral responsibility is untenable. To this the compatibilist can reply 'we have in the intelligent, evolved decision making mechanism something which does all the adaptive work which free will was supposed to do, why not take our notion of responsibility to be referring to it'. An action resulting from the processes of the decision making mechanism in an agent (be they a person, or a mere robot) will likely be reproduced if the agent is in the same or similar circumstances in the future. Undesirable actions will be met with negative reinforcement, desirable actions with positive reinforcements. The fact that we favour verbal ('don't do that!') and physical reinforcement over direct brain surgery merely reflects what 'mental adjustment' tools were and were not available to us during our evolution. This seems to me, at any rate, a plausible evolutionary account for our reactive attitudes but I do not for a moment imagine I have given a conclusive argument for compatibilism – merely shown that one can seemingly adopt compatibilism without incurring any great expense and if one does so the first presumption of the naïve interpretation of identity between will and the conscious experience of will becomes incoherent. The remainder of this paper will be concerned with what the consequences for the second, more commonly held presumption of the veracity of introspection are if we reject the naïve interpretation.

4. Confabulation and the Contents of Consciousness

Confabulation is the phenomena of the occurrence of beliefs in an agent due to gaps in knowledge concerning the subject of those beliefs. Confabulation appears to occur when information presumed to be there by the brain is not, but the mechanisms which should be in place to indicate a deficit are absent. The confabulated belief is the next most probable inference ruling out the possibility of undetected error. Confabulation is usually only brought in to refer to abnormal cases, but it seems to me there are perfectly pedestrian aspects of conscious experience apt to be called confabulations. Because my own application of the concept of confabulation is rather wider than is conventional I shall begin with conventional examples of confabulation and move slowly towards what I see to be the application of the concept in the interpretation of Libet's results and for the nature of consciousness in general.

* Korsakoff's syndrome can result in both retro- and anterograde amnesias. Patients with memory 'gaps' often use environmental clues or inferences to form beliefs in the absence of genuine memories. These confabulatory memories are genuinely believed by the patient. To my mind this is a more extreme, but inherently similar phenomenon to the problems often encountered with the reliability of witness testimony.

* Split-brain patients (those who have had their corpus callosum severed) can be found to confabulate when one hemisphere lacks information the other hemisphere is using to guide action. If you present an image e.g. of scissors in the left visual field and then present a tray of items to pick from, the right hand will – everything going right – choose to pick up the scissors. Ask the patient why they did this however and you will be communicating with the left hemisphere, which will often confabulate an explanation: 'oh, I wanted to cut something out a magazine'.

* Proprioception loss e.g. for a limb will sometimes result in the patient involved suffering bizarre delusions when attempting to explain what the nature of the limb is, excluding the apparently unthinkable – that they have lost their ability to feel the limb. These confabulated explanations range from the not unreasonable 'it's a severed leg someone put in the bed as a prank' to the absurd 'it's my brothers leg'.⁴ Many delusions seem to have this form – inferences made to cover gaps caused by a loss of normal signals from different brain regions.

* Anton-Babinski syndrome is a curious condition in which patients who are blind nonetheless claim to be fully sighted. They will rely on clues about what they should be seeing (they will, if seeing a doctor for example, claim to be able to see a white coat) to formulate their claims about their visual experiences.⁵ My own view of Anton-Babinski syndrome is to take it as another instance of confabulation – beliefs formed to plug gaps caused by the lack visual input in the absence of an internal detector saying 'something is up'.

* Everyday visual experience, on reflection, has aspects of phenomenological confabulation similar to Anton-Babinski syndrome. Because (perhaps) we know the objects of the world are coloured, when they leave the centre of our vision, we insist that we can see colour 'all the way to the edge'. Simple testing with playing cards can show this is not the case; people are no better than chance at detecting the colour of a card in their periphery.⁶ It seems reasonable to take this as an example of common-or-garden confabulation about the purported contents of one's consciousness.

⁴ Examples from Sacks, *The Man Who Mistook His Wife for a Hat*.

⁵ Are these people having phenomenologically similar experiences to fully sighted people. It sounds implausible to say so though as Dennett has shown in 'Quining Qualia' it is not clear that we can separate out claims about the genuine contents of 'ineffable' subjective experience from memories as to what this experience purportedly was.

⁶ Is this a case of 'filling in'. I doubt it; if it were they would need to fill in the space with 'something'; there wouldn't be the same kind of agnosia about the colour of objects in the periphery.

The above cases are all well documented. What is not documented is the manner in which the results from Libet's experiments can be added to this body of evidence against the accuracy of people's conscious introspection. The inaccuracy arising from Libet comes in two forms. First, people believe that their decision to act is simultaneous with their awareness of deciding to act. Not wanting to split hairs, but if we stick with our compatibilist account we can say 'this is not strictly true' because something worth calling volition occurs prior to the conscious awareness of it, but even if we wish to be relatively commonsensical, the reason people are tempted by the naïve interpretation is that it seems so vastly counterintuitive. It is counterintuitive because it seems to imply that the deliberate action and thus *ex hypothesi* all deliberate acts occur unconsciously which is plausible if this is taken to mean simply 'prior to or independent of conscious awareness of' but is flat out incoherent if we expand our notion of what is a 'conscious decision' to what appears to be its normal conceptual scope – all fully deliberative actions. Libet provides us with ammunition to motivate a separation of the executive, deliberative (volitional) faculties and the conscious awareness of these processes which would make Descartes spin in his grave. Secondly, Libet's results also point to the idea that the conscious awareness of volition is a confabulation (of sorts) of whatever mechanism is responsible for providing the kind of internal narrative we associate with consciousness. The conscious experience of volition claims a role as the originator of action to which it is not warranted, which in conjunction with such failures of volitional awareness (such as found in ideomotor and 'alien hand' phenomena) suggest the same kind of confabulation of the powers and scope of consciousness as we have suggested takes place even in common-or-garden visual perception.

What is the significance of this to philosophy in general? Many philosophical positions on mind, particularly dualism, epiphenomenalism and folk psychology but also some adherents of supposedly physicalist positions⁷ rely on the (apparently intuitive) notion that consciousness is a singular, unified, temporally simple⁸ and veridical phenomena: what you think you see is what you get. This view can be attacked by showing either that what you think you are conscious of you are in fact not, or that what you think consciousness is doing, it isn't. The utility of Libet's experimental results is with regards to the latter of these.

6. Conclusions

Free will has nothing to fear from Libet's experiments, so long as we claim nothing more than that free will is the normal operation of frontal lobe regions coordinating action in association with the organism in questions epistemic state to promote its perceived benefit. What does have something to fear is the view that there is something ineffable yet special to be 'found' by looking for consciousness on the basis of our own introspection into the nature of our consciousness. It is yet another fine example of how poor this introspection is, and how little it should be trusted to give a true account of what is going on.

2,632

⁷ I have in mind here John Searle's view that consciousness is a thing 'secreted' by the brain.

⁸ Simply the idea that if I appear to be conscious of an event at time t this is because it is happening at time t and not because it is happening at time t+n and being retroactively indexed to time t.

Bibliography

- Dennett, D.C. 'Quining Qualia' in *Readings in Philosophy and Cognitive Science* Goldman, ed. (MIT, 1993)
- Dennett, D.C. *Consciousness Explained* (Penguin, 1993)
- Dennett, D.C. *Freedom Evolves* (Penguin, 2004)
- Gazzaniga, M. *Cognitive Neuroscience* (WW Norton & Co, 2006)
- Libet, B. 'Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity' in *Brain* 106 (Oxford, 1983) pp 623-642.
- Libet, B. 'Do We Have Free Will' in *The Oxford Handbook of Free Will*, Kane ed. (OUP, 2002)
- Sacks, O. *The Man Who Mistook His Wife for a Hat* (Picador, 1986)