S & CB (2014), **26**, 5–23 0954–4194

PETER G.H. CLARKE

Neuroscientific and psychological attacks on the efficacy of conscious will

Neuroscience and psychology are increasingly being invoked to cast doubt on the fundamental intuition that our intentions and decisions are causally implicated in our behaviour. The initial attack was launched thirty years ago with the famous experiment of Benjamin Libet on the timing of decisions to perform simple movements. A second prong to the attack was launched in 2002 with the publication of social psychologist Daniel Wegner's book, The Illusion of Conscious Will*. I here summarise the intense debate that has resulted and argue that the anti-conscious-will lobby have failed to make an adequate case to justify their iconoclastic claim.*

Key words: Consciousness, will, brain, illusion, neuroscience, Libet, Wegner

The notion that some of our acts are the results of our willing them consciously is fundamental to how we understanding selfhood and human relationships, love and hate, morality and responsibility, sin and repentance. That people do some things on purpose, but not others, seems obvious – and is essential to the very essence of personhood. That this is true of ourselves is our most direct and basic percept. It is undeniable.

Or is it? Over the last thirty years there have been increasing attacks on this very notion, the efficacy of the will. Conscious will[1] is dismissed as an ineffectual epiphenomenon, or even an illusion.

Two different questions: about the will's freedom, and about its efficacy

I am not here referring to the classical debate about the freedom of the will. This is sometimes denied because of the standard argument against free will, which has two parts: 1) if our brains work deterministically, we are not free; and 2) even if our brains work indeterministically, we still don't have free will because our decisions and actions are random. Among modern philosophers, a few hard determinists, including Honderich[2] and Pereboom,[3] deny free will for reasons such as these, but most accept free

1 My use of the term 'conscious' to qualify will may seem redundant, because for most of us the will is by definition conscious, but I use it for emphasis and because those who deny the efficacy of our wills regularly refer to 'conscious will'. The term 'unconscious will' is sometimes used.

2 Honderich, T. *A Theory of Determinism*, Oxford: Oxford University Press (1988).

3 Pereboom, D. *Living Without Free Will*, Cambridge: Cambridge University Press (2001).

will, because they think that the standard argument can be countered in one of two ways. Compatibilists think that the 'varieties of free will worth wanting', to quote Dennett's famous subtitle,[4] are compatible with brain determinism – and they think that brain function is as near as makes no odds to being deterministic. In contrast, libertarians think that free will in the full sense of the term involves a significant degree of indeterminism in the brain – but moderate indeterminism, such as would not produce random behaviour. It would be beyond our present scope to explore the arguments in detail. My reason for mentioning this debate about the will's freedom is simply to make the point that it must be distinguished from the separate debate about the will's efficacy, which is the subject of this essay.

Admittedly, in books and articles attacking free will denials of the freedom and of the efficacy of conscious will are usually linked and are sometimes confused, but the two claims are distinct. My present focus is on the latter claim, the denial of our will's efficacy. It is claimed that our experiences of intending, willing, deciding are merely delayed read-outs, informing us of our brain's decision after the event. Consciousness is not denied, but is claimed not to affect our behaviour.

Defenders of the efficacy of conscious will include both Christians, such as Richard Swinburne[5] and Timothy O'Connor,[6] and atheists such as Daniel Dennett,[7] but most of those who deny it appear to be atheists. These include 'new atheist' propagandist Sam Harris[8] and naturalist philosopher Gregg Caruso.[9] A common feature of the new-style anti-conscious-will lobby is that they invoke neuroscience in two different ways. They attack the freedom of the will by the standard argument, mentioned above; and they attack the efficacy of the will by invoking experimental evidence that is supposed to show that our actions result entirely from unconscious brain processes. The latter attack is grounded mainly on the experiments that neurophysiologist Benjamin Libet initiated in the 1980s[10] and on the arguments and data marshalled by Harvard psychology professor Daniel

4 Dennett, D.C. *Elbow Room: The Varieties of Free Will Worth Wanting*, Oxford: Clarendon Press (1984).

5 Swinburne, R. *Mind, Brain and Free Will*, Oxford, New York: Oxford University Press (2013).

6 O'Connor, T. 'Conscious Willing and the Emerging Sciences of Brain and Behavior', in Murphy N., Ellis G. & O'Connor T. (eds.) *Downward Causation and the Neurobiology of Free Will*, Berlin: Springer (2009).

7 Dennett, D.C. *Freedom Evolves,* New York: Viking (2003).

8 Harris, S. *Free Will,* New York, London etc.: Free Press (2011).

9 Caruso, G.D. *Free Will and Consciousness: A Determinist Account of the Illusion of Free Will*, Maryland: Lexington Books (2012).

10 Libet, B., Gleason, C.A., Wright, E.W. & Pearl, D.K. 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): the unconscious initiation of a freely voluntary act', *Brain* (1983) 106, 623-642; Libet, B. 'Unconscious cerebral initiative and the role of conscious will in voluntary action', *Behav. Brain Sciences* (1985) 8, 529-566.

Wegner in his influential book *The Illusion of Conscious Will*.[11] I shall call the view that is founded on this twin foundation the 'Libet-Wegner thesis'.

In the present paper I argue that both foundations of the Libet-Wegner thesis are unsound, and that the modern denial of the efficacy of conscious will is therefore unfounded.

The brain-mind relationship and the biblical view of humanness

Before addressing the efficacy of conscious will, we need to consider briefly the different ways of conceiving the mind-brain relationship. It is generally accepted that the electrical activity of our brains somehow underlies our conscious thought, including our decision making. How a physical thing, the brain, can be the basis of consciousness is a subject of debate that has given rise to many different philosophical positions, but these can be grouped in two main categories: dualism and monism.

So great was the influence of Descartes on western philosophy that, despite ongoing controversy, from the late seventeenth century until the early or middle twentieth century many or most westerners accepted some form of interactive dualism, involving an immaterial soul acting on a material brain. This view then lost favour, for a variety of reasons including the arguments of secular philosophers such as Ryle, Place, Smart and Feigl in the mid twentieth century. In this same period, interactive dualism lost popularity among Christian academics for the additional reason that advances in the analysis of biblical texts tended to support a holistic conception of man as a unity, not two separate parts in interaction.[12] This realisation was not entirely revolutionary, because there had always been a holistic (or monistic)[13] strand in Christian thought due to the influence of Aristotle on Thomas Aquinas Thus, thomistically inclined Anglican theologian Austin Farrer criticised the interactive dualist views of neurobiologist (and future Nobel prize-winner) John Eccles, writing:

> We will have nothing to do with the fantastic suggestion, that what the supersensitive 'reactors' in the cortex react to, is the initiative of a virtually disembodied soul. To what, then, are we to say that they do react? What else, than to the motions of the embodied soul, that is to say, other motions in the same nervous system?[14]

11 Wegner, D.M. *The Illusion of Conscious Will,* Cambridge, MA: Bradford Books (2002).

12 For review, see Green, J.B. *Body, Soul and Human Life: The Nature of Humanity in the Bible*. Carlisle: Paternoster (2008).

13 Aquinas followed Aristotle in conceiving of the soul as a principle inherent in the body, not a separate immaterial entity outside the body. Some authors describe Aquinas as a dualist, because this word can take on different meanings, but his thought was monistic in my present sense that he thought of the soul as a principle inherent in the body.

14 Farrer, A. *The Freedom of the Will,* London: A & C Black (1958), p. 87.

Among biblical scholars, there is still some debate on the question of body-soul monism or dualism. All agree that the Bible emphasises the unity of human nature and teaches the resurrection of the body, not the Platonic concept of an eternal soul that floats out of the body-prison at death to enjoy an eternity of bodilessness. Many (most?) biblical scholars, including Joel Green, maintain that the dominant biblical view of man is ontological monism with no place for a separate soul.[15] But others, including John Cooper, draw attention to New Testament passages that seem to teach an intermediate state after death during which we will temporarily be disembodied souls while waiting for the resurrection of the body.[16] To emphasise both the unity of human nature and the possibility of the soul to be separate from the body, Cooper uses the term holistic dualism.[17]

Christian philosophers and neuroscientists likewise adopt a range of positions. Many, including most authors of articles in this journal,[18] reject interactive dualism because of the data of neuroscience and the arguments of mainstream philosophy, but avoid also the other extreme of eliminative materialism that rejects mind as illusory. Instead, they mostly adopt an intermediate position, either a non-interactive form of dualism such as emergent dualism,[19] or a moderate monist position such as two-aspect monism (also called dual-perspective theory), according to which our own subjective, first-personal account of our inner life and neuroscience's objective, third-personal account of our brain's activity refer to complementary aspects of a single entity.[20] The first and third-personal accounts are complementary. Thus, my first-personal claim that I performed a movement as a result of a deliberate decision corresponds to (maps onto) a neuroscientist's third-personal report that my movement resulted from the brain activity underlying the decision.

On the other hand, a significant minority of Christian philosophers (and physicists, but not neurobiologists) is resistant to current orthodoxy

15 Green, J. *op. cit.*, (12).

16 Cooper, J.W. *Body, Soul & Life Everlasting*. Grand Rapids MI: Eerdmans (2000).

17 or 'dualistic holism' (in the preface of the 2nd printing (2000) of *Body, Soul & Life Everlasting*).

18 Booth, D. 'Human nature: unitary or fragmented ?', *Science & Christian Belief* (1998) 10(2),145-162; Murphy, N. 'The Problem of Mental Causation: How Does Reason Get its Grip on the Brain?', *Science & Christian Belief* (2002) 14(2), 143-157; Jeeves, M. 'The Boyle Lecture 2008: Psychologising and Neurologising about Religion: Facts, Fallacies and the Future', *Science & Christian Belief* (2009) 21(1), 225-254.

19 Hasker, W. 1999. *The Emergent Self*, Ithaca, New York: Cornell University Press (1999).

20 Nagel, T. *The View From Nowhere,* Oxford: Oxford University Press (1986), chap. III, p. 28; Jeeves, M. & Brown, W.S. *Neuroscience, Psychology and Religion*, West Conshohocken, PA: Templeton (2009).

and continues to support different forms of interactive dualism.[21] These authors are not always explicit about the neuroscientific implications of their positions, but as I understand it most would maintain that the brain activity underlying the act of deciding to move is caused by, or at least influenced by, a nonphysical soul. But, downstream from the neural decision mechanism, the planning and execution of the movement can be assumed to occur by standard neurobiological mechanisms.

Thus, for most monists and dualists, the claim that we do things on purpose, by deliberate acts of will, maps onto the neurobiological claim that the neural activity underlying our conscious decision to act causes, or at least influences, the action. Our central question is whether the anti-conscious-will lobby have succeeded in showing this neurobiological claim to be false.

The Libet claim about the initiation of simple movements

The modern assault on the efficacy of conscious will began with the neurophysiological experiments of Benjamin Libet and his collaborators in the 1980s,[22] which have been interpreted by the authors and by many others as showing that our brains initiate simple voluntary movements before we are consciously aware of the will to move, implying that our conscious will is not involved in the initiation of the movement. I shall refer to this controversial claim as the 'Libet claim' for brevity. It continues to be hotly debated.[23]

The Libet experiment and the mind-brain relationship

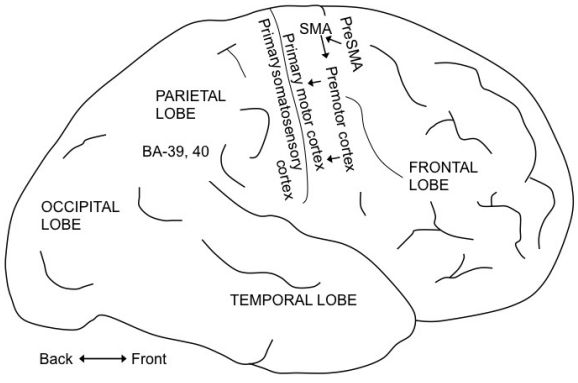
A striking thing about the Libet claim is that it goes against the main versions of both dualism and monism. Interactive dualism predicts that mind events should precede the corresponding brain events, since the mind events are considered to be the real causes, and causes always precede their consequences. Dual-aspect monism (like other forms of monism) predicts that mind and brain events should be synchronous, since mind-level descriptions and brain-level descriptions are considered complementary (and equally valid) accounts of the same processes. But if brain events come before the mind events to which they correspond, this would support

21 Moreland, J.P. & Craig, W.L. *Philosophical Foundations for a Christian Worldview,* IVP Academic (2003), chaps, 11 & 12, The Mind-Body Problem (pp. 222-266); Turl, J. 'Substance Dualism or Body-Soul Duality?' *Science & Christian Belief* (2010) 22, 57-80; Goetz, S. & Taliaferro, C. *A Brief History of the Soul*, Chichester, UK: Wiley-Blackwell (2011); Swinburne, R. *op. cit.*, (5).

22 Libet, B. et al. (1983) *op. cit.*, (10).

23 Sinnott-Armstrong, W. & Nadel, L. (eds.) *Conscious Will and Responsibility: A Tribute to Benjamin Libet*, Oxford: Oxford University Press (2010).

Fig. 1. The brain's cerebral cortex, viewed from the right side. SMA: supplementary motor area. BA: Brodmann's area.



some form of epiphenomenalism, the view that mind events are mere by-products of brain events, with no causal role.

Before going into details about the Libet experiment, I must first provide some information about the neurophysiology of voluntary movement.

The neurophysiology of voluntary movement

It is important to be clear about what is, and is not, being claimed when a movement is called voluntary. Even though these movements involve, by definition, an act of will, that is not to say that every aspect of the movement is conscious or willed. For example, a person engaged in conversation while walking home may be unconscious, or only marginally conscious, of his movements, but they are voluntary in the sense that he intended to walk home. Likewise, the movements of a tennis player as she serves are voluntary, but their control involves many automatic subroutines in the cerebellum and elsewhere. Furthermore, to claim that voluntary movements are caused by conscious acts of will is not to deny that the acts of will arise out of brain processes that are largely unconscious.[24]

What is the nature of the 'I' (or self) that wills the movement and performs it? The use of such terms need not imply interactive dualism. The 'I' (or self, or mind etc.) is conceived by most monists as being embodied in (or emerging from) the brain's activity.

The neural circuits involved in voluntary motor control are exceedingly complicated, and I here give only some simplified information that is necessary for understanding the Libet experiment. Voluntary movements are controlled primarily by the motor cortex (in the back part of the frontal

24 Gomes, G., *Consciousness & Cognition* (1998) 7, 559-595.

lobe – Fig. 1) but in cooperation with many other motor centres including the basal ganglia and the cerebellum. Motor commands are sent from primary motor cortex (and to some extent from other areas) to motoneurons in the brainstem and spinal cord, which in turn control the muscles. The initiation and programming of movements depends on activity in many areas including the supplementary motor area (SMA) and the preSMA, and several areas in the parietal cortex. These areas feed directly or indirectly into the premotor cortex and motor cortex. Electrical stimulation of the motor areas elicits movements, but not usually the urge to move, although stimulation of SMA at intensities too weak to produce a movement has occasionally evoked an irrepressible desire to move going beyond the patient's will.[25] In contrast, electrical stimulation of areas BA-39 and BA-40 in the parietal lobe (Fig. 1) never elicits movements but produces an urge to move, experienced as if it really was the subject's own desire.[26]

The Libet experiment, a challenge to the role of conscious will

An important background to the Libet experiment was the discovery in the 1960s that, before people make a voluntary movement, there is a slow build-up of electrical potential measured from the skull over the motor cortex, beginning as much as a second earlier for simple movements and even longer for complex series of movements.[27] This electrical change is called the readiness potential (RP).

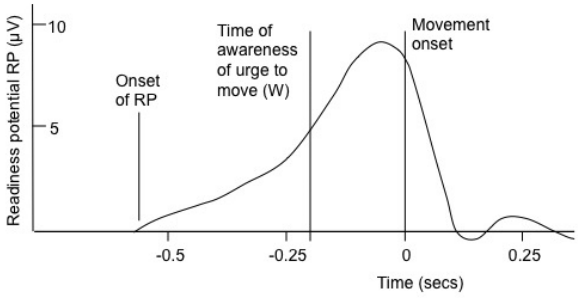
Libet was interested in the relative timing of the RP compared with the movement and the conscious decision to move. He therefore asked his experimental subjects to perform simple movements, in most cases flexion of the fingers or wrist, and to estimate the time of conscious awareness of the urge (or will or decision) to move (W) by reporting the position of a spot moving in a circle on an oscilloscope screen. They were told to perform the movement whenever they felt like doing so, and to pay close attention to the time when they were first aware of the 'urge to move'. He also recorded the RP by electroencephalography, and the time of the movement itself was estimated from the electromyogram. Libet found that time W came only about 200 msec before the movement, whereas the RP began much earlier, usually about 550 msec before the movement (Fig. 2). The fact that the change in brain potential occurred before the conscious decision was interpreted by Libet and by many commentators to imply that our conscious decision to act is not the true cause of the movement. They deduced that conscious will is too slow to make things happen, and that volitional acts must result from unconscious processes in the brain, not

25 Fried, I., Katz, A. et al. 'Functional organization of human supplementary motor cortex studied by electrical stimulation', *Journal of Neuroscience* (1991) 11(11):3656-3666.

26 Desmurget, M., Reilly, K.T. et al., *Science* (2009) 324, 811-813.

27 Kornhuber, H.H. & Deecke, L., *Pflügers Archiv*. (1965) 284, 1–17.

Fig. 2. Schematised readiness potentials (RPs) preceding self-initiated voluntary acts, as in the Libet experiment. Since these scalp-recorded potentials are of only about 10 μV, smaller than the background EEG, the experimenters had to average about 40 raw recordings to obtain reproducible results. We follow Libet et al. in using the term 'Time of awareness of urge to move' and in designating it by W (for will).



from conscious willing. This seemed to imply that our intuitive notion of conscious will must be an illusion.

There appeared to be a small loophole in that Libet's subjects still had the power to veto a movement in the 200 msec between time W and the movement. He therefore argued that even though the initiation of the movement was not the result of conscious will, its vetoing was. This argument was important to Libet himself, who did not wish to deny our fundamental intuitions about consciousness and free will, and has been supported by eminent free-will philosopher Robert Kane[28] and others, but its experimental basis has not been investigated in detail.

The Libet experiment provoked considerable interest and intense controversy, and stimulated further experimentation.

Single neuron recordings during the Libet experiment

It is rarely possible to record from single neurons in the brains of humans, but this can occasionally be done in epilepsy patients using electrodes that have been implanted to localise the zones that cause seizures. Thus, remarkably, Itzhak Fried and his collaborators managed to record from more than 1,000 neurons in the medial frontal cortex of epilepsy patients (and especially in the supplementary motor area, which generates most of the early part of the RP) as they performed the Libet experiment. It was found that a few neurons changed their firing rate (by an increase or a decrease) almost 1.5 secs before time W, and more and more neurons did so over the following 1.5 secs, with about 25% of the neurons firing several

28 Kane, R. *The Significance of Free Will*, New York, Oxford: Oxford University Press (1996), p. 232.

tenths of a second before W. The authors conclude that their findings support the view that the experience of will emerges as the culmination of premotor activity starting several hundreds of msecs before awareness.[29]

Criticisms of the Libet claim

Despite the fame of the Libet experiment and its frequent acceptance in popular and semi-popular writings, it has been the subject of intense controversy. Indeed, most specialists in the philosophy of free will who have addressed the Libet claim have rejected it.[30] Most of the criticisms have focused on difficulties of judging the time of awareness, of interpreting the RP, or of philosophical interpretation, as is discussed below.

Problems of judging the time of awareness

It was central to Libet's claim that the readiness potential began distinctly before time W. The published data of several groups do indeed support this claim, but critics have objected to the reliance on subjective recall after the event to determine time W, because there is evidence that this can be very unreliable. Furthermore, those such as Alfred Mele[31] who have tried the experiment for themselves have found that W is difficult to define. I have done this too, and you may wish to try it using a 'clock' available on Bob Doyle's 'information philosopher' website.[32] When I try this, I find it very hard to judge the precise time when I decided to move my finger/wrist. It would be useful to be able to quantify the reliability of our judgements, but this is difficult to do for a purely subjective decision. For this reason, several research groups have instead measured how reliable people are at judging the time of perceptual events, which is easier to do. Results have been variable and several groups found serious biases that depended, for example, on the modality of the stimulus (visual or auditory or tactile) and on the speed of the clock.[33] This raises doubts about whether one can trust the timing judgements in the Libet experiment. A different critique of the timing was made by Dennett and Kinsbourne,[34] who pointed out

29 Fried, I., Mukamel, R. & Kreiman, G., *Neuron* (2011) 69, 548-562.

30 Bayne, T. 'Libet and the case for free will scepticism', in Swinburne, R. (ed.) *Free Will and Modern Science*, Oxford: Oxford University Press (2011); Mele, A.R. *Effective Intentions: The Power of Conscious Will*, Oxford: Oxford University Press (2009) ; Radder, H. & Meynen G. 'Does the brain 'initiate' freely willed processes? A philosophy of science critique of Libet-type experiments and their interpretation', *Theory & Psychology* (2013) 23, 3-21.

31 Mele, A. R. *op. cit.*, (30).

32 There is a suitable clock at http://www.informationphilosopher.com/freedom/libet_experiments.html

33 Danquah, A.N., Farrell, M.J. & O'Boyle, D.J. 'Biases in the subjective timing of perceptual events: Libet et al. (1983) revisited', *Consciousness & Cognition* (2008) 17, 616-627.

34 Dennett, D.C. & Kinsbourne, M. 'Time and the Observer', *Behav. Brain Sci.* (1992) 15, 183–247.

that Libet's experiment involves an attention shift from the participants' subjective intention to the clock, which may have introduced temporal mismatches between the felt experience of will and the perceived position of the clock hand. A further complication is that transcranial magnetic stimulation applied over the presupplementary motor area after the movement affects the subject's estimation of time W.[35] This shows that the estimation of W depends partly on neural activity occurring after the movement, emphasising still further the difficulty of relying on subjective recall after the event.[36]

To try to solve some of the above problems, Matsuhashi and Hallett devised an alternative methodology for estimating time W. With this, they found that the RP (which they called BP1) occurred before W in only about two thirds of the subjects; worse, the lateralised RP (LRP) that we shall discuss below, always occurred after W.[37]

In view of the controversy about the measurement of subjective timing, considerable attention was devoted by the public media to a paper published in Nature Neuroscience that used brain scanning technology (functional magnetic resonance imaging – fMRI) in a Libet-like experimental paradigm, and included in the summary a claim that a 'decision can be encoded in brain activity of prefrontal and parietal cortex almost 10 seconds before it enters awareness'.[38] After all the subtle debate about a few hundreds of milliseconds, 10 seconds was an enormous time, and the wording of the abstract gave the impression that the temporal priority of the neural decision with respect to the subjective one was finally established. I assume that some journalists and bloggers only had access to the abstract (available free on the web) and not to the full paper, because the main text made only the much weaker claim that the activity of prefrontal and parietal cortex was *correlated with the decision* (to use the left or right hand) with 60% prediction accuracy, up to 10 seconds before the conscious decision. That is very different! To justify confidence in the claim that a neural decision is reflected, the correlation would need to be at or close to 100%, not 60% (which is not very different from chance – 50%). The paper provided intriguing information about brain activity leading up to a decision

35 Lau, H.C., Rogers R.D. & Passingham R.E. 'Manipulating the experienced onset of intention after action execution', *J. Cog. Neurosci.* (2007) 19, 81-90.

36 For a more detailed critical discussion of the validity of Libet's method for measuring the timing of the decision to move, see: Guggisberg, A.G. & Mottaz, A. 'Timing and awareness of movement decisions: does consciousness really come too late?', *Frontiers in Human Neuroscience* (2013) 7, 385, 1.

37 Matsuhashi M., & Hallett M. 'The timing of the conscious intention to move', *Eur. J. Neurosci.* 28, (2008) 2344-2351.

38 Soon, C.S., Brass, M., Heinze, H.J. & Haynes, J.D. 'Unconscious determinants of free decisions in the human brain', *Nature Neurosci.* (2008) 11, 543-545; Bode, S., He, A. H. et al. 'Tracking the unconscious generation of free decisions using ultra-high field fMRI', *PLoS ONE* 6(6):e21612. Epub 2011 Jun 27.

from a surprisingly long time beforehand, but did nothing to rescue the Libet experiment from the criticisms about timing. Another problem is that subsequent research suggests that the early slightly-better-than-chance prediction rate may have resulted from subtle experimental biases.[39]

The overall conclusion on timing has to be that the problems have not been resolved.

Doubts as to whether the readiness potential reflects a decision to move

The Libet claim assumes that the RP reflects a neural 'decision' to move, and that the neural activity underlying the RP causes both the will to move and the movement. Even if such causality could be demonstrated, this would not strictly be sufficient to validate the Libet claim, because the decision must presumably be caused by a chain of pre-decisional neural events, and the RP might reflect some of these. But the Libet claim certainly assumes causality. This is part of the claim, and it has never been proved.

To be precise, we are really talking about the earliest part of the RP, because the timing argument focuses on the RP's onset. To attribute such a decisional and causal role to this earliest part of the RP seems surprising, because it originates mainly in the SMA (Fig. 1), which has been known for more than thirty years to be strongly activated when subjects imagine a complex movement without actually performing it.[40] This is not to deny that activity in SMA can cause movements in some cases, such as when it is stimulated electrically, but it cannot be assumed that the earliest part of the RP necessarily reflects neural processes underlying a decision to move. And there are at least six specific reasons to doubt this.

First, even though electrical stimulation of the SMA can cause movements, it only rarely causes an urge to move, which is evoked much more readily by stimulation of parietal areas.[41] Only in the latter case does the patient feel that the urge was her own, not alien. This suggests that the RP does not cause the will to move.

Secondly, if the RP truly caused the conscious will and the movement, one would expect trial-to-trial variations in the onset of the RP to correlate with trial-to-trial variations in time W; that is to say that trials with an

39 Lages, M. & Jaworska, K. (2012) 'How predictable are "spontaneous decisions" and "hidden intentions"? Comparing classification results based on previous responses with multivariate pattern analysis of fMRI BOLD signals', *Front Psychol.* (2012) 3, 56.

40 Roland, P.E., Larsen, B., Lassen, N.A. & Skinhoj, E. 'Supplementary motor area and other cortical areas in organization of voluntary movements in man', *J. Neurophysiol.* (1980) 43, 118-136.

41 Desmurget et al. *op. cit.*, (26).

early RP should also have an early W. Haggard and Eimer tested this, using a variant of the Libet experiment, and found there was little correlation, ruling out the RP as a cause of the will or decision to move. They did, however, find that the 'lateralized readiness potential' (LRP: i.e. the RP from the cortex on the opposite side relative to the movement minus the RP from the same side) gave a positive correlation, suggesting that the brain processes underlying the LRP might cause the will to move.[42] At the time, their paper did not seem to challenge the Libet claim, because the LRP seemed to fulfil the role formerly attributed to the RP. However, the LRP occurs later than the RP, and subsequent experiments have sometimes found that the LRP occurs even after time W, as is discussed above.[43] So the LRP seems a fragile candidate to replace the RP. Moreover, a very recent paper has failed to repeat the findings of Haggard and Eimer and concludes that both the RP and the LRP reflect brain processes independent of will and consciousness.[44]

Thirdly, Alfred Mele has pointed out a flaw in Libet's experimental paradigm that vitiates attempts to deduce a causal influence between the RP and the movement (and the will to move).[45] In all Libet's experiments, the permanent storage of electroencephalographic data was triggered by the finger/wrist movements. This was necessary as part of the averaging procedure that is necessary to detect the RP, which would otherwise be masked by other concurrent activity in the EEG. If there was no movement, the data were not stored, so any RPs that occurred without being followed by movements would not have been detected. If such RPs without movement did occur, then RPs are not sufficient to cause movements, and more probably reflected brain activity occurring prior to the decision to move. This possibility is difficult to evaluate, because the averaging procedure has to be triggered at a moment defined by the movement.

Fourthly, experiments by Herrmann and others cast further doubt on the interpretation of the RP as causally related to the decision and movement.[46] These researchers used a modified version of the Libet experimental paradigm, in which the participants were instructed to press one of two buttons, depending on a presented stimulus. An RP occurred well before the motor response, as in the Libet experiment. But, importantly, it occurred even before the stimulus presentation, so it clearly did

42 Haggard P. & Eimer M. 'On the relation between brain potentials and the awareness of voluntary movements', *Exp. Brain Res.* (1999) 126, 128-133.

43 Matsuhashi & Hallett *op. cit.*, (37).

44 Schlegel, A., Alexander P. et al. 'Barking up the wrong free: readiness potentials reflect processes independent of conscious will', *Exp. Brain Res.* (2013) 229, 329-335.

45 Mele *op. cit.*, (30).

46 Herrmann, C.S., Pauen, M. Min, B.K., Busch, N.A. & Rieger J.W. 'Analysis of a choice-reaction task yields a new interpretation of Libet's experiments', *Int. J. Psychophysiol.* (2008) 67, 151-157.

not reflect a decision as to which button to press. The authors argue that the RP does not specifically determine the movement, but may reflect a general expectation (which is indeed what the RP was initially thought by Kornhuber and Deecke to reflect, not a decision but a state of readiness, hence its name).

Fifthly, Trevena and Miller devised a modified version of the Libet experiment in which participants made spontaneous decisions to move, or not, and found that the RP was no greater before a decision to move than before a decision not to move, which is not what one would expect if the RP reflected a neural decision to move.[47]

Sixthly, computational analysis combined with some additional experiments suggests that the neural decision to move occurs only very late during the time-course of the RP, not at its onset.[48]

There are thus strong reasons to doubt that the earliest part of the RP reflects neural events underlying a decision to move. This further undermines the Libet claim.

Philosophical implications of the Libet claim

Even if the Libet claim, that our brains initiate movement before we are aware of the will to move, is accepted – which is very controversial, as we have seen – there is also debate about the philosophical implications.

I have systematically used the term 'conscious will' rather than 'free will' because the Libet experiment is more relevant to the efficacy of the will than to its freedom. Nevertheless, many supporters of the Libet claim, including Libet himself,[49] have used the term 'free will' and have attempted to draw implications about human responsibility. This has aroused further controversy, because many critics have argued that Libet's experimental paradigm was irrelevant to the question of free will and responsibility. When we talk about free will, we are usually referring to choices among a variety of options, often with moral implications, and this may require careful deliberation over a period of minutes or hours or days. The Libet experiment is just the opposite. The subject was not making a moral decision, and was not even deciding whether to move, but only when. (The whether decision was resolved by agreeing to take part in the experiment.) Moreover, the subjects were specifically instructed not to deliberate but to act spontaneously, and in their original 1983 paper Libet

47 Trevena, J. & Miller, J. 'Brain preparation before a voluntary action: evidence against unconscious movement initiation', *Consciousness & Cognition* (2010) 19, 447-456.

48 Schurger, A., Sitt, J.D. & Dehaene, S. 'An accumulator model for spontaneous neural activity prior to self-initiated movement', *Proc. Natl. Acad Sci. USA* (2012) 109, E2904-2913.

49 Libet, B. *Mind Time*, Cambridge, MA & London, UK: Harvard University Press (2004).

and his team explicitly pointed out that their conclusions applied only to spontaneous, rapidly performed movements.[50] Thus, even if we ignored the above arguments and accepted that the finger/wrist movements in the Libet experiment were not the result of conscious will, this conclusion could not automatically be extended to situations of human and ethical significance.

For this reason, the anti-conscious-will lobby requires additional evidence at a more cognitive level. Their cognitive-level evidence is discussed below.

Cognitive-level attacks against the efficacy of conscious will

Cognitive psychology and unconscious motivation

Before addressing in detail the cognitive-level attacks against the efficacy of conscious will, it is worth considering the more sober evaluations from experimental psychology. The notion of unconscious motivation was discussed well before the time of Freud, but the modern era of its experimental study is sometimes considered to have been initiated in 1977 by a review by Nisbett and Wilson.[51] These authors argued that people lack insight into their own mental processes on the grounds that they misreport the influences that caused their behaviour, and are reported (but controversially) to be little more accurate in explaining their behaviour than are other people. A more recent review in *Science* emphasises the importance of what the authors called 'unconscious will',[52] and mentions many examples including people's tendency to talk more softly when there is a picture of a library on the wall, or to keep their desk tidier when there is a scent of cleaning agent in the air, without being consciously aware of the influence of their environment.

For reasons such as these and many more, most psychologists consider that motivations and decisions are influenced unconsciously. There is ongoing debate about the importance of the unconscious influences, and a recent review criticises several of the landmark studies on methodological grounds and goes so far as to conclude that 'evidence for the existence of robust unconscious influences on decision making and related behaviors is weak and many of the key research findings… can be plausibly explained without recourse to unconscious influences'.[53]

50 Libet B. et al. 1983 *op. cit.*, (10).

51 Nisbett, R.E. & Wilson T.D. 'Telling more than we can know: Verbal reports on mental processes', *Psychological Review* (1977) 84, 231-259.

52 Custers, R. & Aarts, H. 'The unconscious will: How the pursuit of goals operates outside of conscious awareness', *Science* (2010) 329, 47-50.

53 Newell B.R. & Shanks D.R. 'Unconscious Influences on Decision Making: A Critical Review', *Behavioral and Brain Sciences* (2014, in press).

In short, there is debate about the extent to which our consciously willed decisions and actions may be biased by unconscious processes. Most accept that unconscious bias can be significant. But this is quite different from the extreme claims of the anti-conscious-will lobby, that unconscious processes are all-important and that conscious will is totally inefficacious.

Daniel Wegner's attack on the efficacy of conscious will

Daniel Wegner is a professor of psychology at Harvard and the 2011 winner of the American Psychological Association Award for Distinguished Scientific Contributions. His attack on the efficacy of conscious will is contained in his provocative 2002 book, *The Illusion of Conscious Will,*[54] which draws on a wide range of psychological literature to support his view that conscious will is an illusion. By this he means that conscious will is just a 'feeling' without causal potency, a *post hoc* interpretation. He claims that we do not really make consciously willed choices in any objective sense. He states: 'Both the legal and the religious free will theories assume that the person's experience of conscious will is a direct sensation of the actual causal relation between the person's thought and action'[55] and makes clear that this assumption is the main object of his attack.

In support of his claims, apart from some pages devoted to the Libet experiment, which Wegner finds totally convincing, Wegner's main focus is numerous strange phenomena where notions of self and agency are distorted. These include cases of automatism (such as automatic writing, the movement of Ouija boards, water dowsing, and dissociative experiences), obedience to posthypnotic suggestion, and delusions of 'virtual agency' in multiple personality disorder and in spirit possession and mediumship (interpreted naturalistically). He shows that our drive to consider ourselves as causal agents is so strong that it can lead us to take responsibility for actions that we did not intend, as in cases of patients with the corpus callosum sectioned who confabulate (make up stories) to explain their behaviour that is controlled by the nondominant (usually right) hemisphere that is disconnected from the speech centres. He also reports his own ingenious experiments where subjects are tricked. For example, a person whose arms are hidden under a robe while he watches himself in a mirror can be induced to believe that someone else's moving arms, extending through holes in the robe, are his own and are being moved according to his own volition.

54 Wegner, D.M. *op. cit.*, (11).
55 *ibid.*, p. 336.

Wegner's controversial claims have been debated in great detail by numerous authors,[56] and I can here give only a superficial survey of the debate. A major objection has been that the demonstration of illusions or confabulations in artificial situations or in brain-damaged patients is insufficient to support Wegner's strong claim that conscious will is always illusory even in ordinary situations and in people with undamaged brains. Other frequently expressed general criticisms include the following.

1) Wegner's assumption that the will is just a feeling, ignores the difference between the experience of will and the will.

2) Even though Wegner explicitly rejects Cartesian interactionism, some philosophers consider that he nevertheless makes outdated Cartesian assumptions.[57]

3) His use of the word 'illusion' does not fit even his own conception, because while denying the conscious will's efficacy he maintains that it is often accurate in its indication of cause and effect and is useful as a guide to understanding ourselves.[58]

4) Even if Wegner were correct in his claim that our explanations of our motivations are after-the-fact reconstructions based on incomplete information, this would not justify his use of the loaded word 'confabulation' for normal situations,[59] because our explanations are often accurate. A less common criticism is that of Edward Kelly, who objects that naturalistic interpretations are too readily invoked for paranormal phenomena.[60]

As for the details, the book describes such a wide range of phenomena that it is necessary here to select. I shall focus on just two examples.

Automatisms

The subject of chapter 4, these are movements that appear to be consciously controlled but are not accompanied by a feeling of conscious will. They are

56 Mele, A. R. *op. cit.*, (30); Sharlow, M. F. *Yes, We Have Conscious Will* (2007 Preprint). Retrieved 22 February, 2013 (Available from the PhilSci Archive Web site: http://philsci-archive.pitt.edu/3778/); McClure, J. 'Attributions, causes and actions: Is the consciousness of will a perceptual illusion?', *Theory and Psychology* (2012) 22, 402-419; Nahmias, E. 'Scientific challenges to free will', in O'Connor & Sandis (eds.) *A Companion to Philosophy of Action*, Hoboken NJ: Wiley-Blackwell (2010), chap. 44; a variety of opinions are expressed in: Pockett S., Banks W.P. & Gallagher S. (eds.) *Does Consciousness Cause Behavior?* Cambridge MA: MIT Press (2006), pp.169-186; Carruthers, P. 'The illusion of conscious will', *Synthèse* (2007) 159, 197-213; Morris S.G. 'The impact of neuroscience on the free will debate', *Florida Philosophical Review* (2009) 9, 56-78.

57 Caruso, G.D. *op. cit.*, (9), pp. 210-211.

58 Wegner, D.M. *op. cit.*, (11), pp. 15, 327.

59 *ibid.*, pp. 171-184.

60 Kelly E.F. '[Review of] The Illusion of Conscious Will by D. Wegner', *J. Scient. Explor.* (2003) 17, 166-171.

used by Wegner to support his main thesis of the disconnection between actions and conscious will. As an example, let us consider Wegner's discussion of table-turning in his chapters 1 and 4. This curious phenomenon began in the context of the nineteenth century spiritist movement. People sat around a table with their hands on it, believing (perhaps hoping) that a spirit might intervene, and after some time (which could be almost an hour) the table would begin to move. The cause of this was first studied scientifically by Michael Faraday, who placed force measurement devices between the participants' hands and the table, and found that the source of the movement was their hands.[61] The gentle pressure from the hands of several participants was enough to move the table, even though they were not in the least conscious of contributing to the movement. It appears also that such movements only occurred when the participants were expecting a spirit to act, so the conscious expectation of movement led to the movement via an unconscious mechanism. Wegner provides many such examples of unconsciously mediated movement and argues that this undermines our presuppositions about voluntary control. I find this to be grossly overstated. We all know that our thoughts can bias our movements and our posture without our being aware of it. The thought of a delicious meal awaiting me makes me walk faster without my being aware of my increased speed. I slouch in my chair when disappointed, without meaning to do so. That unconsciously mediated biases exist is widely accepted and unproblematic (even if disputed),[62] but this does not justify the much stronger claim that conscious will plays no role at all.

Protecting the illusion

This is the title of chapter 5, which claims that we think of ourselves as 'ideal agents' who have goals and know consciously what the goals are before we pursue them. Wegner maintains that this is 'all a fabrication, of course' and gives numerous examples of people who invent false stories about their motivations to justify their mistaken belief that they were acting as ideal agents. Wegner gives examples of this happening in ordinary situations and more strikingly in situations of brain damage or hypnotic suggestion. As an example of the latter, he mentions a nineteenth century hypnotist who said to a hypnotised woman: 'After you wake up you will take a book from the table and put it on the bookshelf.' She did just that, and explained her motivation by saying 'I do not like to see things so untidy; the shelf is the place for the book.' This seems to me to fall far short of justifying Wegner's strong claim that conscious will is an illusion. The fact that we invent (or, more often, complete) retrospective explanations of our intentions when they are unknown (in this case) or forgotten (in some other cases) may tell us something about the difficulty of recalling

61 Faraday, M. 'Experimental investigation of table turning', *Atheneum* (July 1953) 801-803.
62 Newell, B.R. & Shanks, D.R. *op. cit.*, (53).

our intentions in special cases such as the ones Wegner mentions. But to extrapolate from this to the claim that conscious will never affects our actions seems to me the height of temerity. To take an analogy, a naive hypothesis might suggest that when we look at an object, say a grid of parallel lines, with only one eye open, we should see a gap in the lines where they go through the 'blind spot', the region of retina where there are no photoreceptors. We don't. Our brains interpret the gap as being due to the blindspot and therefore 'fill in' the gap, and we see the lines as continuous through the blindspot. A Wegner-style visual scientist might be tempted to deduce that our vision is illusory, because the visual image that we see is constructed. Well, yes, the more we learn about vision the more we understand that it does indeed involve a great deal of construction – and our ability to interpret what we see depends on this. Constructed yes, illusory no! Visual illusions do occur, of course, but it would make no sense at all to say that all our vision was illusory. We know also that memory is an active, constructive process and that we fill in gaps to produce a coherent account. I find it unsurprising that we fill in gaps in our memories of our motivations, but this does not refute conscious will.

Thus, even though the debate is still on[63] and my present review is too short to do it full justice, I do not think the arguments of Wegner and his supporters are sufficient to justify his strong claim that conscious will is an illusion. In claiming this, I do not mean to dismiss his idea that the brain mechanisms for being aware of our willing may be separate to some extent from the mechanisms leading to the willed action. This seems to me an interesting open question, whose future elucidation will require some very detailed neuroscientific investigations. These have not yet been performed.

The conceptual framework of the Libet-Wegner thesis

My approach so far has been to take separately the anti-conscious-will arguments based on the Libet experiment, or those based on Wegner's book, and to show that they fail to refute the efficacy of conscious will. I have not attempted to prove that conscious will is efficacious, but consider that the onus of proof is on those who claim to refute the common-sense position.

I now draw attention to some problems with the overall conceptual framework of the Libet-Wegner thesis or of some of its advocates.

Several of the latter seem to have rather obscure notions of the positions that they claim to refute. For example, motor system specialist Patrick Haggard, a former collaborator of Libet and pre-eminent exponent of the Libet claim and its supposed anti-free-will implications, appears to have

63 Caruso, G.D. *op. cit.*, (9).

in mind only rather marginal notions of free will. For example, in a major review published in 2008 on the neuroscience of volition, he mentions the possibility that the brain's circuits might be influenced by 'an unspecified and uncaused cause (the "will")', rejects this view, and concludes the article by stating that 'modern neuroscience is shifting towards a view of voluntary action being based on specific brain processes…'.[64] This gives the impression that 'modern neuroscience' is gradually triumphing against the illusion of free will, but this is confusing for at least two reasons. First, only a tiny minority of modern philosophers conceives of the will as an 'uncaused cause', so why use such a marginal definition? Secondly, the words about modern neuroscience 'shifting towards a view of voluntary action being based on specific brain processes' are strange, because this has been the standard view in neuroscience for over half a century. In the same review, Haggard states that the Libet experiment 'seems to disprove the everyday concept of "free will"'; his reference to 'everyday concept' suggests he recognises that this challenge does not extend to more sophisticated concepts of free will.

An awkward feature of the Libet-Wegner thesis is that it implies a strange semi-epiphenomenalism. Epiphenomalism is the view that our mental phenomena (conscious thought, feelings etc.) are caused by physical events in the brain but have no influence on the brain (or on anything). The Libet-Wegner thesis is in some respects epiphenomenalist and yet their methodology requires the assumption that epiphenomenalism is false, because their case is founded on after-the-event reports about conscious will: either its timing (Libet) or its motivational basis (Wegner). But reporting requires action (muscular movement for speaking or writing or pointing etc.), so their whole approach assumes that conscious will has to be efficacious with respect to the movements involved in reporting even though they claim it is inefficacious for the simple movements (Libet) or complex behaviour (Wegner) that were under study. This curious feature of the Libet-Wegner thesis may not be self-refuting, but is certainly awkward.

Conclusion

Our belief that we do things by deliberate acts of will implies that the neural correlates of our decision to act cause (or at least contribute to) the action. The anti-conscious-will lobby contests this by invoking: 1) the experiments of Benjamin Libet and his successors on the timing of decisions to move, and 2) the diverse phenomena reviewed by Daniel Wegner in his 2002 book *The Illusion of Conscious Will*. The Libet experiments were, however, always difficult to interpret, and new neuroscientific data

64 Haggard, P. 'Human volition: towards a neuroscience of will', *Nature Rev. Neurosci.* (2008) 9, 934-946.

produced over the last decade cast doubt on Libet's central claim that the neural 'decision' to move occurs well before awareness of the will to move. The phenomena reviewed by Daniel Wegner do indeed suggest that our insights into our own agency may not be quite as transparent as we may have thought. But they fall far short of justifying Wegner's extreme claim that conscious will is an illusion. Thus the efficacy of conscious will remains intact.[65]

Dr Peter Clarke works in the Department of Fundamental Neuroscience of the University of Lausanne, Switzerland, where he was an associate professor until his retirement in 2012.

65 The major part of this paper devoted to the Libet experiment is an expanded version of the author's Faraday Paper No. 17 (2013): https://www.st-edmunds.cam.ac.uk/faraday/Papers.php. I am grateful to Martyn Frame and Stuart Judge for their helpful comments on the latter.