Sven Walter
Institute of Cognitive Science
University of Osnabrueck
Albrechtstraße 28 D-49069 Osnabrueck

s.walter@philosophy-online.de

Does Consciousness Cause Behavior?, edited by *Susan Pockett*, *William P. Banks*, and *Shaun Gallagher*. Cambridge, MA: The MIT Press 2006.
ISBN: 0-262-16237-7

According to our commonsensical, manifest image of the world, human beings are freely deliberating conscious agents that behave the way they do *because* they have the beliefs and desires they have. The possibility that the feelings of volition and agency that accompany our behavior may be illusory and our beliefs and desires only ineffective epiphenomena of the brain processes that actually cause our behavior sounds preposterous, to say the least. And yet, scientists have long cast doubt on the assumption that we are the autonomous authors of our behavior that know what they do and why they do what they do.

Back in the late nineteenth century already, Thomas Huxley (1874) famously argued that we are *conscious automata*, comparing consciousness to the steam-whistle which accompanies the work of a locomotive engine but has no causal influence upon it. In the 1980s Benjamin Libet and colleagues discovered that simple motor actions are preceded by a readiness potential in the brain which occurs roughly 350 milliseconds *before* the subject in question becomes conscious of the 'urge' to act, showing that what appears to be a free action, consciously initiated by the subject, is in fact fully determined by prior unconscious brain processes (Libet 1985). More recently, Harvard psychologist Daniel Wegner has argued that the feeling of 'conscious will' that usually accompanies our actions can be present even in cases where the subject does not perform the action, suggesting that the feeling that we have willfully cause an action is an *ex post facto* interpretation by our brain that is as fallible as any other causal interpretation and not at all the reliable indicator for the activity of an authoritative agent or self (Wegner 2002).

Quite often, philosophers interested in the implications of these experimental results have difficulties adequately assessing and interpreting them because they lack an adequate training in the relevant psychology or neuroscience. Conversely, the conclusions neuroscientists, psychologists and researchers from the empirical social sciences draw from their evidence often seem premature from a philosophical point of view. For that reason, *Does Consciousness Cause Behavior?* is an interesting and important addition to the ever growing bulk of literature on consciousness and brain research. According to the editors' introduction, the book "springs from a desire to examine, place in context, and discuss the implications for society of those lines of evidence" (p. 1), and indeed it offers both, a philosophically informed and detailed but for the non-specialist still fairly approachable discussion of the relevant neuroscience, and a range of original and highly interesting philosophical perspectives on its consequences for issues like free will, mental causation, agency, or self-consciousness.

Does Consciousness Cause Behavior? is divided into three parts — 'Neuroscience,' 'Philosophy,' and 'Law and Public Policy' — and brings together sixteen essays (including one reprint), by biologists, cognitive scientists, neuroscientists, law scholars, philosophers, and psychologists.

Part one primarily deals with the exact temporal order of and the interrelations between the neurophysiological correlates of conscious acts of intention on the one and the initiation and control of the corresponding actions on the other hand. In line with Libet's original results, Susan Pockett argues that in the case of simple motor actions conscious volitions arise only *after* the neural processes eventual-

ly resulting in the movement have already begun. Going beyond Libet, she then argues that the monitoring and correction of ongoing movements is also accomplished by the brain (by means of an 'efference copy mechanism' which automatically and unconsciously compares intentions with peripheral feedback and which, as Suparna Choudhury and Sarah-Jayne Blakemore suggest in their contribution, is also used to distinguish our own actions from those of others). Things may be different in the case of *complex decisions* and *long-term intentions* (usually formed in the dorsolateral prefrontal cortex and the presupplementary motor area) and the actions ensuing from them (usually initiated somewhere in the frontal cortex/basal ganglia loops). Here, the problem is that "while it is clear that consciousness is generally *associated with* these processes, nobody has yet been able to design experiments that would unequivocally nail down the temporal relationship between the appearance of this consciousness and the onset of whatever neural events underpin the intentions and movement initiations' " (p. 22). This corroborates a point often made in philosophical discussions of Libet's experiments: since simple motor actions like pressing a button or moving a finger are quite unlike the long planned actions we usually perform, it is problematic to conclude from Libet's experiments that the latter, too, are initiated unconsciously by the brain. Unless these experiments can be extended to complex actions, and Pockett's paper nicely explains why at least at present this is impossible, their implications for the debate about free will are limited.

Libet instructed subjects to perform a simple motor activity within a certain time frame at an arbitrary moment decided by them and to remember the earliest moment (called 'W') at which they were aware of the 'urge' to act by noticing the position of a dot circling a clock face (the 'clock' actually being a cathode ray oscilloscope modified so as to be able to measure extremely short time intervals). W was found to occur on average roughly 200 milliseconds before the movement, but roughly 350 milliseconds *after* a readiness potential that eventually resulted in the movement was measurable. On one interpretation this shows that the actions in question are not the result of the subject's conscious intention to act because consciousness just comes too late to play a causal role, and if our actions are not the result of our conscious intentions to act, free will would seem to be an illusion. Libet himself resisted this line of reasoning, suggesting instead that we can retain some degree of freedom because after W there still remain approximately 200 milliseconds for consciousness to 'veto' the execution of the action. This 'veto account' of free will, however, turned out to be problematic for various reasons. First, a study by Haggard and Eimer in 1999 showed that Libet's results cannot only be obtained for simple go/no go-tasks (press a button/don't press a button), but also for choice-tasks (flex your left wrist/flex your right wrist). While Libet claims that we first consciously decide *what to do* before unconscious brain processes 'decide' *when* to perform the action and consciousness finally has its veto-option, Haggard and Eimer's study arguably shows that not even the decision *what to do* is made consciously. Second, in *Does Consciousness Cause Behavior?*, Richard Passingham and Hakwan Lau cite new experimental evidence which suggests that "the demand to attend may bias the temporal judgments to be too early" (p. 58), so that the subjective experience of the 'urge to act' may in fact occur *much later* than Libet assumed. The subject's judgment about the occurrence of W can be influenced by a transcranial magnetic brain stimulation which occurs up to 200 milliseconds *after* the action, suggesting that W may not occur 200 milliseconds *before* the action, but up to 200 milliseconds *afterwards* and then be 'backdated,' so that consciousness could not even exert any kind of veto since the 'urge to act' doesn't become conscious until the action is already over.

A more philosophical discussion of Libet's

experiments can be found in part two, where Alfred Mele questions the assumption that the readiness potential can be identified as the neural substrate of an *intention* or *decision* to act. While Libet indifferently talks about an 'urge' to act and an 'intention' to act, Mele shows that a lot hinges on which of these descriptions is appropriate: If Libet's point is merely that we cannot consciously control our *urge* to act, then that wouldn't seem to conflict with our intuitions about free will, for no one would claim that we can enjoy free will only if we can consciously initiate all of our urges. What would be a problem is if our *intentions* to act could be initiated unconsciously, but Mele argues convincingly that there is nothing in Libet's experiments that warrants such a conclusion; drawing on empirical evidence from a reaction time study by Haggard and Magno (1999), he defends an alternative interpretation according to which the onset of the readiness potential corresponds to a relatively unspecific desire or urge to act which is then followed later, at time W, by the conscious intention to act (and it is this conscious intention, and not the unconscious urge occurring earlier, that causes the action).

The perhaps most interesting question addressed in part two is what exactly cognitive science can contribute to the debate about free will. Peter Ross, for instance, argues that since the controversy between compatibilists and incompatibilists is grounded in a purely semantic quarrel about the proper understanding of the notion of 'control,' no empirical discovery can ever resolve it. However, he maintains, science can resolve the debate between libertarians on the one and those who hold that indeterminacy cannot be sufficient for free will on the other hand, for instance by showing that the kind of quantum indeterminacy typically alluded to by libertarians does not exist. Ross admits that this would leave untouched libertarian accounts like that of Timothy O'Connor which do not appeal to quantum indeterminacies, but he points out that O'Connor's account is undermined by scientific findings like those of Nisbett and Wilson in the seventies or, more recently, Wegner to the extent that the assumption that *introspection* is a reliable epistemological tool is questionable.

Although Ross doesn't explain why the failure of introspection in the artificial scenarios of Nisbett and Wilson and their likes should support the claim that it is impossible to base one's account of free will on the assumption that introspection is *by and large* a reliable epistemological tool, he does draw attention to a very important point which is too often ignored in current debates about free will. As long as *compatibilism* about free will can be made plausible, neuroscientific evidence to the extent that our actions are in fact determined by brain processes do not threaten free will, given that the point of compatibilism is precisely that determinism is compatible with our having the kind of control over our actions that makes them count as free. However, this compatibility alone does not show that we are in fact free and that the cognitive sciences cannot threaten our freedom. The fact (if it is a fact) that determinism is compatible with our having some kind of control over our actions solves the problem of free will only if we in fact *do enjoy that kind of control over our actions*, while recent findings from cognitive science seem to show that many of our actions are *not* under (the right kind of) conscious control. For instance, Wegner argues that we do not have reliable first-person access to the fact that we have willfully performed an action because we can be lured into thinking that we did what in fact someone else did or that we did not do something we actually did. If in order to have control over our actions we have to be conscious of the fact that we perform them, and if free will is indeed possible only if we have that kind of control over our actions, then the experiments Wegner appeals to seem to undermine free will even if compatibilism is true.

Here, I think, lies the real threat for our hope for free will, and not in the neuroscientific evidence presented by Libet *et al.*, and one laudable thing about *Does Conscious*

Cause Behavior? is that many of its contributions take up this pivotal issue, most of them critical of Wegner's work. Timothy Bayne argues that the *general reliability*, or veridicality, of our feeling of consciously authoring our own actions is left untouched by Wegner's research, and Elisabeth Pacherie, who is primarily interested in a dynamic theory of intentions and not so much in the problem of free will or mental causation, correctly points out that to show that conscious will is *sometimes* an illusion is not to show that it is *always* illusory. Yet, how good a response this is depends upon how widespread the phenomena Wegner appeals to are, and it seems that they are much more common than usually assumed (see, e.g., Wilson 2004).

One possible solution would be an account of free will like the one defended by Shaun Gallagher who argues that the answer to the question *Does Consciousness Cause Behavior?* has no bearing on the issue of free will. Even if actions are initiated unconsciously in the brain already before we are conscious of our intention to act, this does not entail that we are 'unfree' because free will is a longer-term phenomenon and thus independent of proximate motor initiation or motor control. According to Gallagher, free will "involves temporally extended deliberative consciousness that is best described as a situated reflection" (p. 121) and is entirely compatible with the fact that the proximal causes of movements happen at a sub-personal, unconscious level. Assuming that long-term situated reflection is always conscious, an account like Gallagher's would solve the problem discussed above.

If such an account turns out to be unavailable, there are a couple of questions that must be addressed by any serious discussion of the problem of free will and that have as of yet not received the attention they deserve. Must we be conscious of the fact that we are acting for reasons and must we be conscious of the reasons for which we are acting in order for us to be free? Do empirical studies like those appealed to by Wegner or Wilson show that we fail to be conscious of our reasons for acting to a significant extent? If so, does this undermine our general capacity for free actions, or only the *degree* to which our actions are free? I don't have answers to these questions, and neither do the contributors to *Does Consciousness Cause Behavior?*, but we need answers if we want to make progress on the problem of free will.

Part three on law and public policy is the weakest and by far least interesting. Leonard Kaplan from the University of Wisconsin's Law School argues that "neuroscience has such a powerful rhetorical strength that its claims will be likely to have impact beyond what it purports to prove" (p. 277), suggesting that the results of neuroscience, the attention it receives in popular media and its "institutional rationalization ... into policy" (p. 298) will eventually result in a shifting understanding of fundamental ascriptions of responsibility and be misused for social control purposes, thereby undermining human autonomy and dignity. The German psychologist Wolfgang Prinz defends a constructivist account of free will according to which free will is not a fact of nature, but a *social construct* resting on intuitions that are shared and communicated among individuals and that "emerge if and when individuals learn, in social discourse, to develop a self as source of action-decisions and actions" (p. 269). Sabine Maasen offers a discourse analytic examination of a debate about the impossibility of free will staged recently in German newspapers, the main protagonists of which were Wolf Singer, Gerhard Roth and, incidentally, Wolfgang Prinz. Drawing on the work of the German philosopher Peter Bieri, Maasen defends a compatibilist account of free will and argues that the arguments by Singer, Roth or Prinz make sense only against the background of a couple of misconceptions which "are not only highly implausible within the framework of everyday experience [but] also based on an 'adventurous metaphysics"' (p. 348) — a perfect example for the mistaken assumption, criticized above, that free will is immune against objections from cogniti-

ve science once the philosophical lesson of compatibilism is appreciated.

The undoubtedly most interesting contribution to part three is Susan Hurley's paper, although it has little to do with the topic of the book. Having reviewed compelling evidence for the claim that viewing media violence leads to an increased tendency towards aggressive behavior, Hurley argues that there is a strong human predisposition to imitate novel actions observed in others (even in cases where they clearly are inappropriate means of obtaining a goal), that imitation is a largely unconscious process, and that an increased tendency towards aggression after exposure to media violence is the consequence of imitation. Viewing media violence therefore not only causes an increased tendency towards violent behavior, it does so unconsciously, bypassing autonomous deliberative processes. From this, Hurley concludes that media violence should not be tolerated and that it cannot even be supported by an appeal to the principle of the freedom of speech.

Does Consciousness Cause Behavior? is an interesting and valuable book for philosophers not blessed with an adequate training in empirical sciences, and the papers in part two on philosophy are original, highly stimulating and relevant even to experts in the field. For those familiar with the neurophysiology and not primarily interested in its philosophical implications, there will not be much news, I suspect, since the contributions to part one, although informative for colleagues from neighboring fields, hardly contain any cutting edge news for insiders and sometimes simply seem to summarize previously published material. And the majority of those interested in the question whether consciousness causes behavior will definitely be able to live without the part on law and public policy — although I highly recommend Hurley's paper to anyone.

References

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

Haggard, P. & Magno, E. (1999). Localising awareness of action with transcranial magnetic stimulation. *Experimental Brain Research* 127, 102-107.

Huxley, T.H. (1874). On the hypothesis that animals are automata, and its history. *Fortnightly Review* 22, 555-580.

Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences* 8, 529-539.

Wegner, D. (2002). *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.

Wilson, T. (2004). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. New York: Belknap.