

Free Will and Modern Science

RICHARD SWINBURNE (Ed)

Oxford, Oxford University Press, 2011

xviii + 206 pp., ISBN 978 0 19 726489 8, £17.99, \$35.00 (paperback)

To be published in *International Studies in the Philosophy of Science*.

This is the author's copy which may differ in minor details from the final print version.

Recently, there has been a resurgence of interest in free will. The present volume collects papers on both traditional philosophical and neuroscientific research on free will, with some suggestions on how these two strands of inquiry could or should interact with each other. It consists of thirteen contributions by established scholars and researchers, including a foreword by Peter Simons and an introduction by Richard Swinburne. Six of the chapters were presented at a British Academy symposium on the topic in 2010, and the remaining papers were written for this volume. The contributions are a mixed bag. Some aim to make an original contribution, some provide overviews and summarize familiar arguments, and some are largely based on the author's past work. Overall, I think that this volume makes a good contribution to the free will debate, albeit partly for the purpose of providing introduction and overview. In what follows, I will give a brief summary of each chapter. My commentary will in some cases be very brief and in some more extensive, depending on how much interest (or disagreement) the chapter has sparked in me.

In the first chapter, Patrick Haggard provides an overview to recent work in cognitive neuroscience on voluntary action. He distinguishes between two features of freely willed actions and explains what neuroscience can tell us about them. The first feature is that we "could have done otherwise". Haggard assumes that this ability is compatible with determinism, and he argues that neuroscience shows that we are able to do otherwise in the sense that the process of action selection is a process of 'progressive reduction' at the beginning of which multiple actions are neurologically possible (16). I found this interesting, but not particularly convincing, for Haggard seems to mean the selection of a specific movement when he talks about the selection of action. When we

say that we could have done otherwise, we mean usually that we could have performed a different *action*, not just a different *movement*. This is important, for one and the same action may be performed with different movements. According to Haggard, the second component of free will is a sense of agency. Here, Haggard summarizes Libet's well-known work on action initiation and some more recent findings in support of the view that the sense of agency is not merely a post-act interpretation. Like much of the scientific work on voluntary action, the approach here is based on the distinction between *externally triggered* and *internally generated* actions, where freely willed actions are identified with the latter. This approach is methodologically well-motivated, but it is potentially misleading when it is used to capture free will. No one should think that freely willed actions must be generated exclusively by internal states. To the contrary, freely willed actions should be reason-responsive, and as such they should be responsive to the relevant external events.

The next chapter by Tim Bayne focuses on Libet's challenge to free will. There is a large amount of literature on this. Bayne provides a lucid and helpful reconstruction of the challenge and he argues, as many before him, that the Libet experiment does not show that free will is an illusion. As part of this, he suggests that subjects in this experiment experience their decisions as 'points of origins of the action' (34) and that there is, therefore, a serious danger that their experience is illusory. I am not convinced, however, by this interpretation. Subjects are instructed to perform a certain movement when they become aware of an urge to do so. Given this, it seems that they are responding to an internal event in accordance with the instructions. I fail to see why we should think that this is accompanied by an experience of being a 'point of origin' (whatever that means, exactly).

Frank Jackson explores the implications of physicalism for free will. First, if type physicalism is true, then it does not matter whether or not determinism is true, because chancy elements cannot enhance an agent's control. This line of argument is familiar. But it is also controversial, and Jackson does not have anything to say about this controversy. Then he argues that anomalous token physicalism has no significant implications either. On this view, our actions are free from determination, as there are no strict psychological

laws. However, this does not matter, Jackson argues, because the view assumes that our *movements* are fully determined.

In his own contribution, Richard Swinburne argues that it is most unlikely that neuroscience will ever be able to predict all human actions. His argument is based on familiar *a priori* considerations from the philosophy of mind (about privileged access, the holism of the mental, and so on). According to Swinburne, neuroscientists would need to know an ‘enormously large number of detailed laws’ (79), and it is unlikely that this knowledge can ever be attained. Given this, he argues, we have reason to assume that our actions are not determined, ‘in the absence of counter-evidence’ (82). However, had Swinburne consulted recent neuroscience, he would have discovered that neuroscience does not need any laws to make claims about prediction. Neuroscience can infer the existence of predictive brain events by means of pattern recognition methods (usually on the basis of fMRI data). Further, given what we already know about the neural correlates of cognitive capacities, one might just as well hold the opposite of Swinburne’s claim: in the absence of counter-evidence, we have reason to assume that intentions are causally determined (or, at least, predictable), just like all other mental states that are realized by the brain.

Harald Atmanspacher and Stefan Rotter address the question of whether the brain is a deterministic system. They remind us that we cannot draw metaphysical conclusions from ‘epistemic descriptions’ (86), and they show that the question of determinism in the brain is complex. We must distinguish between different levels of description (sub-cellular, neuronal, assembly of neurons), and reasons to assume (in)determinism at one level do not necessarily amount to reasons to assume (in)determinism at the other levels. Their conclusion is that the current state of neurobiology does not settle the question of whether the brain is a deterministic or an indeterministic system.

Solomon Feferman objects to the argument for indeterminism (in ‘mathematical thought’) from Gödel’s incompleteness theorems. The basic idea here is that Gödel’s theorems seem to show that mathematical thinking cannot be ‘mechanical’, as they show that there are mathematical truths that cannot be proven by formal systems. In his reply to Feferman, J.R. Lucas defends this line of argument. Whatever the truth on this matter, it should be noted that this is of relevance to the free will debate only if incompatibilism is

true. It is assumed that the falsity of mechanistic theories of the mind would ‘leave open the possibility of making real choices’ (120). But, of course, if free will is compatible with determinism, it is also compatible with mechanistic theories of the mind.

Galen Strawson’s chapter is on his argument for the impossibility of moral responsibility, which has been published in several versions elsewhere (1986 and 1994, for instance). This argument has been widely criticized and Randolph Clarke has even offered a diagnosis of why ‘Strawson’s argument has impressed so few’ (2005: 13). Strawson does not address any of the criticisms, and he contributed a chapter that is virtually identical to the 1994 article.

In the next chapter, Helen Steward argues that free will is incompatible with determinism because human agency is incompatible with it. She assumes that agents are ‘entities that things can be *up to*’ (145), and in a footnote she argues that this does not beg the question. I think she is right about this. But there is a more fundamental worry here. Compatibilists argue that *free* agency can be part of a deterministic world. Incompatibilists deny this, but usually they grant that agency can be part of a deterministic world. If it is assumed that agency cannot be part of a deterministic world, then it follows, trivially, that free agency is incompatible with determinism. And so it seems that this approach begs the question. Steward might respond that she does not simply assume that agency is incompatible with determinism, but that she argues for it. If determinism is true, she argues, then an agent is a mere ‘place or vessel’ where events take place—then the agent is not *doing* anything (146). However, I think that this is just non-causalist rhetoric. Moreover, this position faces not only the usual compatibilist rejoinders, but it faces also the arguments for the causal theory of action (which are widely accepted). So, far from strengthening the case for libertarianism, I think that this approach weakens it. Nevertheless, the idea is interesting and this chapter is the most original contribution to this volume.

Howard Robinson’s chapter offers a defense of substance dualism based on considerations concerning personal identity. This is an interesting argument, but, as Howard acknowledges, ‘almost everything’ that he says in defense of the view is ‘highly controversial’ (176). For our purposes, the important question is whether there are any significant implications for the free will problem. Howard suggests that property dualism

is too weak for libertarian freedom—libertarianism requires substance dualism. However, many contemporary libertarians would dispute that. I think the main issues, not addressed by Howard, are the following here: if there is an irreducible self, does this self cause actions in virtue of having certain properties? If so, how is this different from property dualism? If not, how is the self's causality different from randomness?

The final chapter by R.A. Duff is on criminal responsibility. Duff provides a clear and helpful introduction to the criminal law and he asks whether a skeptical argument concerning the impossibility of 'ultimate' responsibility (such as Galen Strawson's) would undermine criminal responsibility. Duff argues, plausibly, that it would not, simply because criminal responsibility does not presuppose 'ultimate' responsibility, but only compatibilist reason-responsiveness. It is difficult to specify conditions for an agent's reason-responsiveness. Duff suggests that an agent can be shown to be reason-responsive if there is a coherent narrative according to which the agent is 'in principle persuadable' (190)—in my opinion, an interesting and promising suggestion.

References

- Clarke, R. 2005. On an argument for the impossibility of moral responsibility. *Midwest Studies in Philosophy* 26: 13–24.
- Strawson, G. 1986. *Freedom and Belief*. Oxford: Clarendon Press.
- . 1994. The impossibility of moral responsibility. *Philosophical Studies* 75: 5–24.

MARKUS SCHLOSSER
Department of Philosophy
University of Leiden
PO Box 9515, 2300 RA Leiden
E-mail: m.schlosser@hum.leidenuniv.nl