SWINBURNE ON THE CONDITIONS FOR FREE WILL AND MORAL RESPONSIBILITY

DAVID P. HUNT

Whittier College

Richard Swinburne's rich and stimulating *Mind, Brain, and Free Will* synthesizes data and theories from diverse areas of philosophy into a single unified account of what agents are and how agency is exercised. The resulting account is currently unfashionable, but that says more about fashions than it does about the viability of Swinburne's account. Since I'm largely sympathetic to the picture Swinburne paints, I will offer friendly criticisms of some details of the landscape, focusing on those that most closely concern free will.

The first part of the book argues for positions in ontology, epistemology, and philosophy of mind that lay a foundation for Swinburne's account of agency. Many of these arguments are well worth engaging, but I will simply summarize here the conclusions that are most relevant to this review. (1) Substance causation, not event causation, is fundamental. 'The regularities constituting or underlying laws of nature are regularities, not of ... events, but ... in the causal powers and liabilities ... of actual substances.' (p. 7) Swinburne calls this the 'substances-powers-and-liabilities account', or 'SPL'. (2) We are essentially simple mental substances. Human beings 'consist of two parts – [the] soul (the essential part) and [the] body (a nonessential part), each of them separate substances' (p. 170). Swinburne argues here for 'the simple view' on which 'personal identity is a separate feature of the world' rather than 'analyzable in terms of degrees of continuity' (p. 150). '[E]ach person has a "thisness" which makes him or her that person, a "thisness" other than any thisness possessed by the matter of their brains', so that 'being that person is compatible with having any particular mental properties or any physical properties (and so body) at all' (p. 151). (3) It follows

from (1) and (2) that we exercise a species of causation reflective of the kind of substance we are. 'Such a substance could exercise its power to cause some effect because it intends – that is, tries – to cause that effect, and not because it has a propensity to cause the effect.' (p. 133) This is 'intentional', as opposed to 'inanimate', causation; the more common term is 'agent causation'.

This is of course a highly controversial mix, but the individual parts gain plausibility through their connection to the whole. Agent causation, for example, occupies a much better position against its critics if SPL is correct. I might add that while Swinburne does not situate this book within the Christian theism for which he's argued so powerfully elsewhere,[1] a fair amount of mutual support could be had if he did. Taking agent causation again as an example, the original causal powers belonging to God are presumably agent- rather than event-causal; a theist, then, can either *add* event causation to this original agent causation, or employ agent causation as a model for all causal explanation.[2] It's clear which of these is simpler.

Human agents in fact stand in intimate causal relation to bodies, most intimately to brains. This raises the question how the intentional causality exercised by minds is related to the inanimate causality exercised by the material substances with which they are conjoined, in particular, by those constituting their brains. It is only when the agent can make a difference to the outcome – when an exercise of agent causation brings about a different result (a different intention) than would have been produced by material causes alone – that the agent has free will and qualifies as morally responsible for his or her action.

What are the conditions under which this may happen? It turns out that they are fairly limited. Swinburne has some interesting things to say about how mind and brain might interact on these occasions, but I will take up the question on the psychological level. We are motivated to act by our desires and our value beliefs, 'beliefs about the objective intrinsic goodness or badness of doing actions of different kinds' (p. 175). ('Moral beliefs' are a species of value beliefs that concern 'a special kind of overriding goodness' and 'overlap substantially with the beliefs of most

[1] The word 'God' does not even appear in the book's index.

[2] For Swinburne, agent causation provides this model *because agents are substances*. A different but equally simple approach is offered by occasionalism, on which the only causation *is* agent causation.

other humans' – p. 176.) The problem is that '[b]eliefs are by their very nature involuntary' (p. 77), as are desires (p. 85). But the most common configurations of beliefs and desires make it 'inevitable' that we intend as we do. These are cases in which our intentions

(1) 'are caused by a strongest desire and we have no contrary moral belief',
(2) are caused 'by a strongest moral belief when we have no contrary desire', or
(3) 'simply execute (in the way which we believe to be the quickest way) some ultimate intention.'

In contrast, there are just two circumstances in which the outcome is *not* inevitable:

(A) where 'we have equally strong competing desires and moral beliefs', or
(B) 'where the desires and moral beliefs are in opposition to each other' (p. 201).

In A, where my beliefs and desires are tied in strength, 'I will have to make an arbitrary decision', though this 'will be fully rational, for whatever I do, I have a reason to do it, and no better reason for not doing it' (p. 184). It's only in B that the conditions exist for an exercise of full-on free choice. 'Here I have to decide whether to yield to desire and do the less good action, or to force myself – contrary to my strongest desire – to do the best action. ... This situation I will call the situation of *difficult moral decision*.' (p. 184)

It's worth comparing Swinburne's position with the very similar one endorsed by Peter van Inwagen.[3] Van Inwagen identifies three sets of circumstances in which we can't choose otherwise:

(i) when inclination is unopposed by duty;
(ii) when duty is unopposed by inclination; and
(iii) when it's obvious what to do.

These clearly parallel 1-3 above. Van Inwagen also identifies three circumstances in which we *can* choose otherwise, all involving conflicting alternatives where it isn't obvious what to do:

[3] Peter Van Inwagen, 'When Is the Will Free?' in *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, ed. Timothy O'Connor (Oxford: Oxford University Press, 1995), pp. 219-238.

(a) Buridan's ass cases;
(b) when duty conflicts with inclination; and
(c) when conflicting values are incommensurable.

His (a) and (b) clearly parallel Swinburne's A and B, but Swinburne does not make explicit room for cases like (c). These might be shoehorned into A, on the grounds that incommensurable competing moral beliefs are, if not 'equally strong', at least *not measurably unequal in strength*. But because these cases often elicit a profound wrestling with one's deepest values, they don't really belong with cases that can be settled with a coin toss.

Van Inwagen's argument has elicited responses from a number of critics, including John Fischer & Mark Ravizza.[4] Critiques of Van Inwagen's restrictions on the exercise of free will are equally critiques of Swinburne's. I will make two comments on Swinburne's position, focusing just on the role of beliefs.

First, I'm not persuaded that beliefs are as immune to the will as Swinburne supposes. He argues that the involuntariness of belief (unlike that of thought and desire) 'is a logical matter, not a contingent feature of our psychology' (p. 77). The key step in the argument is this: 'But then [if we thought it was up to us whether or not to believe that *p*] we would know that we had no reason to believe that our belief that *p* was in any way sensitive to whether or not *p* was true; and in that case we couldn't really believe it.' (p. 77) But the fact that people actually do this is the best evidence that it is possible; it's then the job of a good theory to account for this possibility. It's hard to account for it on the assumption 'that a belief that some proposition has an (epistemic) probability on the believer's evidence greater than ½ ... is logically equivalent to a belief in that proposition ... ' (p. 76). There are people – perhaps even many people – who believe not just 'in' God (which may be a different use of 'belief'), but believe to be true the proposition *that God exists* and who would demur if asked whether they believe that the probability of this proposition is greater than ½ on their total evidence. William James famously argued that there are conditions under which such belief is not only possible but epistemically permissible. James' will to believe and Kierkegaard's leap of faith are especially likely to take as their objects

[4] John Martin Fischer & Mark Ravizza, 'When the Will Is Free', in O'Connor, op. cit., pp. 239-269.

value beliefs, the beliefs whose involuntariness is supposed by Swinburne to play such an important role in limiting our free will.

Second, the effect of beliefs on our intentions, like our having the beliefs in the first place, may be more subject to the will than Swinburne allows. Of this effect he writes:

> I could not believe that some action was really morally good to do ... and yet not see myself as having a reason for doing it. And I could not see myself as having a reason for doing it unless I had some inclination to do it. And the better I believe some good action to be, the greater as such is my inclination to do it. (p. 178)

But what seems to have fascinated Augustine about his theft of pears is that it didn't fit this paradigm. Perhaps Swinburne means to block this response with his very broad notion of a value belief, so that when Augustine (or Raskolnikoff) acts contrary to a moral belief that he has precisely because it is a moral belief, he does so in the service of a non-moral value belief that he has. The question is whether there is good independent evidence that such a non-moral value belief is present, or whether its presence is posited simply as a requirement of the theory.

Leaving to one side the frequency of free agency among human beings, let us turn to Swinburne's thoughts on an important challenge to free will, as he understands it. This is Harry Frankfurt's famous critique of the Principle of Alternate Possibilities, or 'PAP', which Swinburne formulates as follows:

> A does *x* freely only if he could have not done *x* (i.e. could have refrained from doing *x*). (p. 203)

(PAP, as it has been discussed in the literature, is actually a principle about moral responsibility; what Swinburne calls 'PAP' is really a freedom version of PAP.) Frankfurt's counterexample involved a case in which Jones decides to kill Smith and both of the following are true: (i) if Black fails to detect a prior sign that Jones will decide to kill Smith, Black intervenes to *cause* Jones to decide to kill Smith; (ii) Black does detect the sign and does not intervene, so Jones decides on his own. Frankfurt believed that in this scenario Jones is morally responsible for his decision while having no accessible alternative to the decision, so PAP is false.[5]

[5] Harry Frankfurt, 'Alternate Possibilities and Moral Responsibility', *Journal of Philosophy*, 66 (December 1969), 829-39.

Swinburne endorses a response to Frankfurt, called the 'dilemma defence', that is associated with David Widerker.[6] Either the prior sign causally determines Jones's decision, or it doesn't. If the former, Frankfurt's scenario begs the question against the libertarian; if the latter, it remains possible for Jones to refrain from deciding to kill Smith – not ultimately, of course (Black will ensure that Jones makes the decision he wishes him to make), but *at t*, the time at which Jones actually decides to kill Smith. All that's needed to save PAP from Frankfurt counterexamples, then, is to clarify it with the help of some temporal indexing:

A does *x* freely at t only if he could have done not-*x* at t instead.

This principle, which Swinburne calls 'PAP*', is 'surely true' (p. 204).

I think that Swinburne's confidence in PAP* is misplaced. Alternatives can always be found in Frankfurt cases; the question is whether they are sufficiently 'robust' to ground all the responsibility ascriptions we wish to make. If Jones can refrain *at t* from deciding to kill Smith, this might explain how he can be morally responsible for deciding *at t* to kill Smith, but not how he can be morally responsible for deciding to kill Smith *full stop* (as he surely is), since he has no alternative to this.[7] But the dilemma defence has also given rise to new counterexamples which violate PAP* as well as PAP. Some feature 'buffered' scenarios in which the agent must complete an intermediate step (traverse a psychological buffer, so to speak) before he is in a position to decide otherwise.[8] Widerker himself, ironically, no longer supports the dilemma defence, and has developed his own counterexample to PAP, which he calls 'Brain-Malfunction-W'.[9]

But there is a vast literature on Frankfurt counterexamples to PAP, and it's unreasonable to expect Swinburne's brief remarks to do it justice.

---

[6] David Widerker, 'Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities', *Philosophical Review*, 104 (1995), 247–61. A similar move can also be found in Robert Kane, *The Significance of Free Will* (New York: Oxford University Press, 1996), and in Carl Ginet, 'In Defense of the Principle of Alternative Possibilities: Why I Don't Find Frankfurt's Arguments Convincing', *Philosophical Perspectives*, 10 (1996), 403–17.

[7] See David Hunt and Seth Shabo, 'Frankfurt Cases and the (In)significance of Timing', *Philosophical Studies*, 164 (March 2012), 1-24.

[8] David Hunt, 'Moral Responsibility and Buffered Alternatives', *Midwest Studies in Philosophy*, 29 (2005), 126–145.

[9] David Widerker, 'Frankfurt-Friendly Libertarianism', in Robert Kane, ed., *The Oxford Handbook of Free Will*, 2nd edition (Oxford: Oxford University Press, 2011), pp. 266-287.

The point I would like to make is not that Swinburne fails to settle the debate in PAP's favour, but that he doesn't need to do so. If the alternatives requirement for free will is abandoned, there is still the sourcehood requirement, for which (like the alternatives requirement) there are both compatibilist and incompatibilist interpretations.[10] For an agent-causal libertarian like Swinburne, it's the sourcehood requirement that is fundamental anyway, and it's not surprising that Swinburne's own definition of 'free will' – 'the agent acts intentionally without their intentions being fully determined by prior causes' (p. 202) – is a pure statement of incompatibilist sourcehood, unsullied by any reference to alternatives. If PAP is false, alternative possibilities (like the red spots signifying measles) are a symptom that ordinarily accompanies free will, though they are metaphysically distinct from it, as shown in the extraordinary cases constituting Frankfurt counterexamples. The underlying condition, of which alternatives are normally symptomatic, is simply a particular kind of causation, the 'intentional causation' exercised by agents, when this is genuinely effective (i.e., makes a difference not explained by inanimate causation alone). This might be a fairly frequent occurrence. Swinburne's limited conditions for free will rest on two premises: that most intendings are inevitable, given the agent's beliefs and desires; and that inevitable intendings – intendings for which the agent has no accessible alternatives – are unfree. I have already suggested that the first premise is too strict, but if PAP is rejected, this argument for limited free will can be resisted at the second premise as well. I think this is a result that Swinburne should welcome.

Free will is important because it's a requirement for moral responsibility, the subject of the book's last chapter. What is the scope of our moral responsibility? To answer this question, Swinburne draws on moral intuition (of course), but also on theories developed earlier in the book. An example is his judgment that the mere passage of time, no matter how much the agent has changed, does not diminish responsibility (p. 226). This is said to follow from his 'simple' account of personal identity, according to which a mental substance has the same essential properties and 'thisness' throughout life; changes in contingent properties, such as memory and character, would not then affect the individual's responsibility. This is not an issue on which ordinary judgment speaks with a single voice, and Swinburne explicitly contrasts

[10] Widerker now characterizes himself as a 'source incompatibilist'.

his position with that of Locke. In the face of someone who protests, 'But I'm not the same person I was 20 years ago!' Swinburne is in effect responding, 'There is (perhaps) a sense in which this is true, and a sense in which it is false; unfortunately, the sense in which it is false is the sense relevant to moral responsibility.' Insofar as one's moral intuitions line up with Swinburne's rather than with Locke's, this may provide some retroactive support for the account of personal identity from which this result is supposed to follow.

I propose to review four further areas in which Swinburne's conclusions about moral responsibility are at least controversial. (The areas are interconnected, so distinguishing among them is somewhat artificial.) The first of these concerns the conditions for praiseworthiness (there are of course companion conditions for blameworthiness). Swinburne's initial claim is that, *normally*, an agent is to be praised only for actions believed by the agent to be morally good but not morally obligatory. The point of the qualifier is soon evident, because praise is also clearly relative to another standard reflecting our expectations of other people and how difficult it was for them to perform the action under assessment (p. 212). It turns out, then, that an agent may be praised for an action that is morally obligatory if it is sufficiently difficult for the agent to choose (given opposing desires), and not praised for an action that is morally good but not obligatory if it doesn't require any extra-normal effort.

I have a couple of worries about this account. One is that the account is complex, combining the initial claim (that an agent is to be praised only for actions believed by the agent to be morally good but not morally obligatory) with a set of exceptions, where the exceptions could just as well have constituted the norm and the norm the exceptions. An alternative account is simply that a person is praiseworthy for doing what (they believe) is morally good – whether or not they believe it to be morally obligatory – when they could have acted otherwise but didn't. In this context, a third party who exclaimed 'Good for you!' would not be saying anything false; but since praiseworthiness comes in degrees, it is often not worth pointing out, and the 'conversational implicature' of doing so would be misleading. This account is simpler and more unified, and for that reason seems to me to be better.

The other worry is that degree of praiseworthiness does not always track degree of difficulty, and sometimes it even seems inversely proportionate to it. I'm inclined to think that a woman who rushes into a burning building to save her child, without stopping to think about it, is

more praiseworthy than a woman who does so only after wrestling with a 'difficult moral decision'. Swinburne later discusses the case of a 'hero who is caused inevitably ... to do a supererogatory action', and argues that we should not praise (or praise so highly) the hero's action; rather we should admire the character from which the action flows, and praise any earlier actions that led to the development of that character (p. 220). This seems to me overly restrictive of what the hero can be praised for, but it also makes it hard to understand how God can be praiseworthy for anything he does. Perhaps divine praiseworthiness rests on a wholly different analysis than human praiseworthiness, or 'praising' God is just a loose manner of speaking (whose real content is admiration for God's character). But a theory on which God is genuinely praiseworthy, in a sense that is continuous with the sense in which humans are praiseworthy, is surely preferable.

The other controversial issues I would like to mention concern what we are culpable *for*. Swinburne's earlier argument that free will is restricted to a fairly rare set of circumstances leads to similar restrictions on moral responsibility. Here are three such restrictions.

We are culpable only for what we do freely at the time we do it. Most of those who restrict free will do not similarly restrict moral responsibility. They are able to do this because they endorse so-called 'tracing' principles, under which a person can be held responsible for an action they didn't freely choose if the action follows from an earlier action which they did choose. (There are different accounts of how the one action must 'follow from' the earlier action; an epistemic condition will surely be part of the mix.) But Swinburne rejects tracing in favour of the strict view that a person whose own choices led to their being unable to fulfil an obligation 'may be culpable – not for the failure to fulfil the obligation at the later time, but for allowing themselves to get into a situation where they believed that it would be improbable that they would be able to fulfil the obligation' (p. 213). But most people's judgment is that when a drunk driver kills a pedestrian, the driver is culpable not only for earlier actions that could have been avoided (such as drinking to excess), but also for actions that the driver was then too impaired to avoid (such as killing the pedestrian). This requires tracing.[11]

[11] Manuel Vargas, for example, writes that 'one of the nice features about tracing is that it is one of a few things to which nearly all parties in the debate about free will appeal to with equal enthusiasm'. 'The Trouble with Tracing', *Midwest Studies in Philosophy*, 29.1 (2005), 270.

We are culpable only for trying, not for the success of our efforts. An excellent justification for this position is that, once we make our contribution to events by trying, the matter is then out of our hands: it's up to the world, not to us, whether our efforts succeed. How can we be blamed for *that*? But when Swinburne concludes that '[s]omeone is just as culpable for trying to blow up a civilian aircraft although prevented from doing so by the police discovering the bomb, as they are for succeeding in blowing up the aircraft' (p. 211), I find my intuitions putting up some resistance. The successful bomber certainly seems to have more on his conscience, more *for* which he is culpable. The same is true of the drunk who kills someone in comparison with the drunk who manages to get home without hitting anyone, each of whom (as we've just seen) is no more nor less culpable than the other by Swinburne's lights. In the case of the successful bomber as in the case of the impaired driver, a tracing principle might produce results more in keeping with ordinary judgment. The two cases also raise the question of 'moral luck', and what a good theory should do with this problem. The failed bomber is the beneficiary of moral luck: through no credit to himself, he has been spared responsibility for multiple deaths. The question is whether such luck is a real feature of the moral landscape, a tragic concomitant of the human condition, or an illusion to be dispelled by the right moral theory. Swinburne's restriction on culpability to cases of trying clearly belongs in the latter camp; whether that's a virtue or a vice is a question I'll leave open.[12]

We are culpable only for acting contrary to our value beliefs. As we saw earlier, Swinburne regards our beliefs at any particular time as givens; for this reason, when we act in accordance with our value beliefs (beliefs whose very nature is to motivate action), we cannot be blameworthy for so acting. We can be culpable then only when acting contrary to our value beliefs. Swinburne offers the following example of culpability being limited by the agent's value beliefs: 'Some people believe that stealing from the rich is not wrong; and so if I have this belief and also the belief that you are rich, I would not be culpable for stealing from you.' (p. 211) Whether or not this is the right result in this particular case, the principle seems too strong. It might, for example, justify the conclusion that Hitler's culpability was much more restricted than anyone would have thought.

[12] The term 'moral luck' was introduced by Bernard Williams in his 'Moral Luck', *Proceedings of the Aristotelian Society*, supp. Vol. 50 (1976), 115-35.

If we learned that Hitler's choices conformed very closely to his value beliefs, this would not (and should not) lead us to a significantly different assessment of his culpability. Perhaps he was obligated to form better beliefs; perhaps he was obligated to act contrary to his value beliefs. But in one way or another, Hitler's culpability needs to be tied less tightly to his value beliefs. Another way to arrive at the same moral is to imagine a parent who takes to heart Kant's dictum that the only thing that is good without qualification is a good will; persuaded that culpability is relative to one's value beliefs, the parent sets about instilling in the child those value beliefs to which it is easiest, and thus most likely, that the child's choices will conform. Culpability-avoidance is not the only quality that parents should try to foster in their children, but something has surely gone wrong if this quality is easier to achieve the more lax the child's value beliefs.

In conclusion, it should be evident that I find human choice and the moral responsibility that comes with it more mysterious than Swinburne makes it out to be. Perhaps Swinburne does as well. Peter van Inwagen, at a key point in his *Essay on Free Will*, summed up the situation as follows: 'I must choose between the puzzling and the inconceivable. I choose the puzzling.'[13] Free will *is* puzzling; to cite the title of a later article by Van Inwagen, it 'remains a mystery'.[14] But philosophers aren't content with this situation; it's our job to work on the puzzle and make the mystery somewhat less opaque. This requires, by the very nature of the enterprise, abstracting from reality what is amenable to philosophical analysis. This Swinburne does with great skill, and the resulting book is an impressive example of how philosophical order can be imposed on the messy phenomenon of human free will.

[13] Peter van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983), p. 150.

[14] Peter van Inwagen, 'Free Will Remains a Mystery', in Robert Kane (ed.), *The Oxford Handbook of Free Will*, 1st edition (Oxford: Oxford University Press, 2002), pp. 158-77.