Review of

The Evolution of the Soul

by

Richard Swinburne

Overview. In The Evolution of the Soul (henceforth TES) Richard Swinburne (henceforth RS) presents his case for substance dualism. In the first part of the book, RS proposes that the most basic components of the mental life are sensations, thoughts, purposings, desires, and beliefs. He then argues that none of these components can be adequately described as physical events (whether outside the body or inside the body), and therefore that ‘hard’ materialism fails. In the second part of the book, RS presents a variety of thought experiments which indicate that, regardless of how exhaustive our knowledge of the physical world is, there are further facts about persons which elude us. This, RS concludes, indicates that persons are composed of a material component (‘body’) and an immaterial component (‘soul’). RS also argues that the most basic ground we have for judging personal identity is, in the end, one rooted in the soul. RS also presents some tentative thoughts on how the soul might have arisen in the course of evolution. The second part of the book thus functions as both an argument against ‘soft’ materialism and an argument for substance dualism. In the third part of the book, RS explores various aspects of the human soul, including the interrelationships among language, rationality, and choice, moral awareness, the structure of the soul, and the future of the soul.

Part I: The Mental Life

Terms. For the purposes of this review the five most important terms are substances, properties, events, mental, and physical. A substance is simply a component of the world which has a temporal history and causally interacts with other substances. A property is a characteristic or a trait of a substance. An event is a substance having, gaining, or losing a property at some time t. A mental property is some property to which an individual has privileged access, and a physical property is some property which is (or can be) equally accessed by all individuals. To make these (possibly opaque) terms more concrete, consider the following example:

At 7:15 AM EST on May 22, 2014, I am having the thought ‘What a wonderful book The Evolution of the Soul is’. While having that thought, I am also wearing a pair of brown loafers.

Let’s now apply our terms to the above example. The substance is me—I am a component of the world, I have a history through time, and that history can be described in terms of causal interaction with other components. I have two properties—a mental one (the thought) and a physical one (wearing brown loafers). The thought is a mental property because I have privileged access to that thought—that is, I have a way of being aware of the content of the thought that no other person has. Put loosely, the thought is ‘naked before my mind’s eye’ in a way that is not true of other persons. My wearing brown loafers is a physical property in that any appropriately placed observer can also know that I am wearing them—I do not have ‘privileged’ access to that property in the way I do the thought. If we specify the substance (‘me’), the time (7:15 AM EST on May 22, 2014) and a property (either the mental property or the physical property) we have isolated and adequately described an event.

Epistemic principles. RS then discusses several ‘epistemic’ principles (most or all of which will be ‘old hat’ to readers familiar with Swinburne’s writings). These principles are credulity, testimony, simplicity, and charity. In order, the principle of credulity (PC) asserts that one should trust a ‘seeming’, in the absence of counterevidence. The PC entails the principle of testimony (PT), which says that one should trust the testimony of others regarding events (again, in the absence of counterevidence). The principle of simplicity (PS) asserts that, all else being equal, the simplest explanation is the most probable. The principle of charity (PC) is entailed by the PS and says that (in the absence of counterevidence) we should assume that the mental lives of others are similar to our own.
Identity criterion. With the terms defined and the epistemic principles in place, RS turns to what it means for two events to be identical (let’s call this the identity criterion). For event A to be the same event as event B, the same substances, properties, and times must be involved. RS then uses this identity criterion to evaluate three different views on the mind/body problem. First is hard materialism. Hard materialism is a position which requires strict identity between physical and mental events. That is, it asserts that the identity criterion is fulfilled in all three facets—whatever the prima facie distinction between physical and mental events is, it is in fact illusory. The same properties occur in the same substances at the same time. Next is soft materialism. Soft materialism asserts that mental properties are not the same as physical properties, but that the mental and physical properties occur at the same time in the same substance. Put differently, humans are exclusively material beings (so there is only one ‘substance’, our bodies) but that our physical substance has both mental and physical properties. Finally, there is substance dualism, which asserts that physical events cannot be the same as mental events because the properties are different and they occur to different substances. On the substance dualism view, physical properties occur in physical substances and mental properties occur in a mental (immaterial) substance—the soul or mind. RS argues against hard materialism in Part I of the book and against soft materialism/for substance dualism in Part II of the book.

Hard materialism rejected. On hard materialism, physical events and mental events are the same events if the same substances, properties, and times are involved. There are two main varieties of hard materialism in the philosophical literature: behaviorism and brain state identity theory. On the former, mental events are identical to public behaviors. In other words, sensations like ‘pain’ are really nothing more than utterances related to pain, shouting ‘ouch!’, etc. On the latter, the sensation of pain is nothing more than a particular brain state (‘c-fibers firing’). In arguing that sensations of pain cannot be identical with physical events outside the body (as on behaviorism) or inside the body (as on brain state identity theory), it would appear that sensations are not physical events at all. Therefore, hard materialism should be rejected. RS rejects behaviorism by observing that individual S and individual S* can both be in pain and S cries out while S* does not. On behaviorism S is in pain while S* does not. But both are experiencing the sensation of pain. Therefore, behaviorism fails. In rejecting brain state identity theory, RS draws upon the work of Saul Kripke. In essence, RS argues that the brain-state identity theorist has confused ‘intrinsic’ descriptions of events with ‘extrinsic’ descriptions of events. The former are what is ‘internal’ and truly essential to some phenomenon (whether an event or a property) and the latter include both the essentials of some phenomenon along with its contingent characteristics (say, causes and effects). For example, an intrinsic description of Brutus stabbing Caesar would be just that: that Brutus stabbed Caesar. An extrinsic description of Brutus stabbing Caesar would be Brutus killing Caesar. The stabbing caused the killing, but it need not have. Similarly, ‘c-fibers firing’ may cause the sensation of pain but it is not identical to the sensation of pain. Put differently, for RS the brain-identity state theorist is confusing the necessary aspects of a mental property or event with the contingent aspects of a mental property or event. One may cause the other, but that is not the same as being the same property or event. RS also provides reasons to think that sensations have real effects. The same general approach is used to reject hard materialism for thoughts, purposings, desires, and beliefs.

Part II: The Soul

Rejecting soft materialism, accepting substance dualism. Having rejected hard materialism, RS turns to arguing against soft materialism and for substance dualism. He does this primarily by invoking thought experiments related to split-brain surgeries. He sketches a scenario in which some person S is told that their brain matter will be divided among two subsequent persons (S* and S**). The first will lead a life of unimaginable luxury and fulfillment, the second one of unimaginable agony, frustration, and torment. Should S be hopeful, fearful, or what? The answer to that question revolves around what happens to S. RS’s point is that we can know what happens to all of the bits of brain matter, and yet not know what happens to S himself. That is, there is a fact of the matter (does S wink out of existence? Is S now S* or S**?) which is not explicable in terms of what happens to the physical bits. We can only know the answer to that question, RS posits, if we know what happens to the immaterial substance
which is the essence of $S$—namely, his soul or mind. Thus there are facts about the world which soft materialism cannot explain, but substance dualism can.\textsuperscript{vi}

**Personal identity and the soul.** RS provides a further (although subsidiary) argument for the acceptance of the existence of the soul by examining how we ground judgments of personal identity. (That is, how we judge that some person $S$ and $S^*$ are either the same or different people. As noted above, his thought experiments invoking split-brain patients are meant to show that it is soul continuity which grounds such judgments.) RS begins by making a distinction between ‘peripheral’ indicators of personal identity and more ‘central’ indicators of personal identity. Peripheral indicators include things like physical appearance, finger prints, bodily continuity, and brain continuity. We judge that persons $S$ and $S^*$ are likely to be the same person to the extent that they appear similar to each other, have similar finger prints, and appear to occupy the same body. We do this, RS believes, because all of these are indicative of brain continuity. Still, why think that brain continuity is an indicator of personal identity? Because, RS argues, brain continuity is itself an indicator of the most central criterion of personal identity—continuity of memory.\textsuperscript{vii} If RS is right, this means that the physical (contingent) indicators of personal identity presuppose the existence of the necessary (mental or ‘soulish’) indicator of personal identity.

**The origin and life of the soul.** RS finally begins to discuss the interplay between the gradual evolution of man from more primitive organisms and his theory of substance dualism. He notes that the tight link between having a mental life and having a certain level of complexity in brain structure means that we are justified in assuming that a fetus has a soul only when a certain level of development is achieved (RS suggests around 20 weeks\textsuperscript{viii}). This link also means that as one ‘descends’ the biological chain from more to less complex organisms, the less justified we become in positing that there is a soul attached to the body of a given organism.

**Part III: The Human Soul**

**The four ‘marks’ of the human soul.** RS then spends four chapters discussing the distinctive facts of the human soul: the use of language, moral awareness, free will, and structure. In discussing the use of language, RS explores the role that language played in our evolution from being able to grasp only very basic concepts to being able to understand complicated cause and effect relationships, relationships of abstract logic, the distinction between necessary and contingency, and our development of moral awareness. His discussion of moral awareness revolves around us having ‘particular’ views of morality (it is good that I and my family should have enough to eat) and, through rational reflection, developing more ‘generalist’ views of morality (it is good that other beings—beyond myself and my family—should have enough to eat). The discussion of free will is fascinating and includes a penetrating discussion of some of the problems with believing in determinism\textsuperscript{ix}, and the discussion of the ‘structure’ of the soul involves a detailed analysis of how human beliefs are tightly interconnected with one another.\textsuperscript{x}

**The future of the soul.** In the final chapter, RS notes that so far he has only argued for the bare possibility that the soul can survive without the body. As a Christian theist, RS does affirm a belief in the immortality\textsuperscript{xii} of the soul and the afterlife. However, he doesn’t think that there is a simple, straightforward philosophical argument for the immortality of the soul which has been supplied in TES. Rather, he feels that

“The theist has first to argue for the existence of God, a person….of infinite power, goodness, and freedom. He may argue that the existence of God provides the simplest explanation of the universe, its virtual total regularity of behavior in conformity to natural laws, and more particular phenomena within the universe. It would then follow that God, being omnipotent, would have the power to give to souls life after death….The Christian theist will further need to show that God intends to bring souls to function after death. He could show this either by showing that it was an obligation on an omnipotent being to do such a thing and that, being good, God would do it; or by showing that God had announced his intention to do so.”—Location 4052, Kindle edition

**Summary**
In Part I of the book, RS argued that mental properties are distinct from physical properties and that only confusing necessary and contingent details of mental properties and events gives hard materialism whatever initial plausibility it holds. In Part II, RS argued that knowing only the physical facts about the world results in an incomplete description. Only if one adds knowledge of immaterial substances (‘souls’ or ‘minds’) can one adequately describe reality. In Part III, RS describes some aspects of the human soul which distinguish it from animal souls.

All in all, TES is a great (and convincing) read. I do have two minor qualms, however. The most basic is this—there is a lot of discussion of the soul and fairly little about evolution per se. Given the title, I expected more detail about the interplay between the two. Second is that it is not clear to me which position RS takes on the creation of souls—has God ‘frontloaded’ the laws of nature so that, given a certain level of complexity, the soul ‘arises’ or is the creation of each soul an act of ‘special creation’? These are, however, minor quibbles. In addition, comparing RS’s views on these matters with those in his more recent Mind, Brain and Free Will (see Footnote ix) should provide for an interesting glimpse into the developing thought of a prominent Christian philosopher.

---

1 Read as ‘intentions’. It appears that RS uses this term to avoid confusion with intentionality (‘aboutness’).

2 RS provides technical definitions of these terms—in fact, he spends a good deal of time fleshing out very precisely what he does (and doesn’t) mean by them. For sake of brevity, I ignore those discussions.

2 For example, it seems that I see a table in front of me. That ‘seeming’ is prima facie justification for believing that I do in fact have thoughts (in the absence of counterevidence). This is defeasible, of course—it is possible that sufficient counterevidence could be brought to bear. If the PC is not embraced, then knowledge can never ‘get off the ground’—as denying the PC entails an infinite regress.

4 The invocation of Kripke involves an extended discussion of what it is for something to be a ‘rigid designator’ (or a ‘name’) of some X and what it is for something to be a ‘description’ of X. While illuminating, this is a bit of a rabbit trail for a book review. Roughly speaking, ‘name’ is parallel with the term ‘intrinsic description’ used in the body of the review and ‘description’ is parallel with the term ‘extrinsic description’ used in the body of the review.

5 As always with RS, there are fascinating details buried in footnotes and sub-paragraphs. The essence of his argument for the effectuality of sensations is that (a) it seems they have effects and so, in accordance with the PC, we should assume they do (b) there is psychological research (based on the imagery findings of Kosslyn et. al) which are best explained by sensations having effects. I am trained as an experimental psychologist, so I found this overlap between my professional and personal interests fascinating (c) we can observe connections between our behavior and apparent sensations and, by PS, invoke the actual existence of sensations as the explanation. RS makes it quite clear that (c) is in one sense an abnormal way to go about providing evidence for the efficaciousness of sensations. Why do that, when we have privileged access to their existence? Nonetheless, he includes this is a possible justification for believing in their efficaciousness. I’m reminded of this quote from John Searle:

“The best comment about behaviorism is the old joke about the two behaviorists after they just had sex. He says to her, “It was great for you, how was it for me?” (Laughter) If behaviorism were right, that ought to make perfectly good sense, because there’s nothing going on in him except his behavior, and she’s in a better position to observe his behavior than he is.” from http://reason.com/archives/2000/02/01/reality-principles-an-intervie/2

6 RS also argues, in a brief appendix, that the existence of immaterial substances (souls) is part of a cumulative case for theism. (See Chapter 9 of his The Existence of God.)

6 For the role that memory continuity can play in a philosophical analysis of heaven, see Footnote 12 of my review of Jerry Wall’s Heaven: The Logic of Eternal Joy at http://www.apologetics315.com/2013/10/review-heaven-logic.
viii I am uncertain of the implications of this for holding a ‘pro-life’ (to use the language of U.S. political discourse) position. One could assume that the soul is simply not expressed until that level of development, although it is in some sense already extant.

ix The argument against determinism involves the ‘counter-suggestibility’ of humans and is as follows. The argument can be phrased in terms of an if...then statement. IF a counter-suggestible human S were informed of the predictions made by a deterministic account of what simple basic action S would take at some future time t, then S would try falsify the prediction (i.e., do some other *non*-predicted basic action) at that future time t. S would be able to do so, as this is a simple basic action. For determinism to survive (not be falsified), the 'if' part of the preceding sentence must not be realizable. That is, it for some reason must be *impossible* for S to *understand* the prediction conveyed to him by the scientist. However, it is most implausible that the deterministic brain state-behavior links would have such a ‘perverse’ character. That is, human brains are quite capable of perceiving information in a wide variety of circumstances, via quite different informational schemes (smoke signals, words, hand gestures, written text) etc. Yet in this type of scenario—again, in any scenario which fits the broad but skeletal schema above, varying in the simple behaviors predicted, the participants, ways of 'getting across' the prediction--somehow the subject becomes incapable of understanding the information. Again, the 'if' (the subject understands the prediction) cannot be realized. Hence we have a very powerful (perhaps overwhelmingly powerful) inductive argument from other areas of human decision making that this unique, perverse, counter-intuitive pattern does *not* hold for attempts to falsify determinism.

While I find the argument fascinating and prima facie persuasive, I am a bit hesitant in endorsing this argument for a rather good reason—RS himself has become dissatisfied with the argument. In personal correspondence, he confirmed that I had correctly understood his argument but also said he has come to harbor some strong (but very technical) reservations regarding it. He indicated that while the rest of his material in TES is consistent with that of his newest book Mind, Brain, and Free Will (or MBFW) he feels that this argument for free will has been superseded by those of MBFW. For an excellent review of MBFW, see http://www.apologetics315.com/2013/05/review-mind-brain-and-free-will-by.html

x The basic gist of the soul structure chapter is this—absorbed by their physical life. RS also notes that the more ‘integrated’ (internally consistent) one’s mental life is, the more difficult it is to change (whether through therapy, surgery, medication, or through force of will). This has interesting implications for our character in the afterlife as well, which RS only hints at.

xi He does, however, explicitly reject Cartesian arguments for the *innate* immortality of the soul.