

**Review of Saunders, G. (2014). *Acts of consciousness: A social psychology standpoint*. Cambridge: Cambridge University Press.**

'Others think of me therefore I exist' (p. 93)

This book has a disarming combination of modesty and grand ambition. It's an unusual read. It's informed equally by philosophy and the arts. It doesn't just talk about experience, but is obviously shaped by specific experiences, and conveys their emotional reality to the reader. Its author, psychologist Guy Saunders, speaks in the first person, and uses barely any jargon. Its first words are 'the most important three and a half words' the author knows, and they appear 15 more times at important points; they are 'I don't know'. I find myself responding to these features as though they were radical, but they aren't, or shouldn't be.

This book is about what it is to be a person. This means it has a lot to say about concepts like consciousness, self, and agency. It approaches them from an unfamiliar direction: the experiences of those who are held hostage, including in solitary confinement. Imprisonment and solitary confinement, Saunders argues, turn down the volume on the everyday, foregrounding what normally seems background. Since many things in ordinary life distract us from our psychological makeup, solitary confinement can clarify what is usually opaque. This logic is similar to the rationale for incorporating mindfulness practice into the study of consciousness: that the clamour of normal life makes us prey to illusions (things that seem other than they are). It also resembles the logic for studying people with brain damage: finding out about how intact brains work by studying what is changed by the damage. Saunders investigates freedom by studying what seems to be its opposite, and his conclusion from the study of prolonged solitude is that prisoners survive solitary confinement by creating social worlds: we exist only by virtue of other people.

Freedom and consciousness are linked by the idea that consciousness (what Saunders calls *minded consciousness*) means knowing this experience could have been otherwise, and freedom means I could have done otherwise. Both are active capacities not passive possessions or occurrences: they are ways of knowing (consciously) and acting (freely). To test out the parallel, Saunders reimagines Frank Jackson's thought experiment about Mary the colour scientist as Mary the captive scientist: does she learn anything new about how it feels to be free when she gets out of the room where she learnt all there was to know about freedom? If our intuition says yes, what does this tell us about the case of conscious colour vision? Saunders tries out many ways of pumping our intuitions and uncovering them, from well-known thought experiments to practical tasks like drawing the outlines of spaces between objects in a scene. All these methods help attune our attention to what is usually background: I can draw a stick me, but I can't draw a line around most of what makes me a person – my relationships with others. (Hence Derek Parfit's famous teleporter wouldn't work unless you could take your entire place-time with you.) From philosophical thought experiments and visual aesthetics come complements to the conclusion from Saunders's analysis of his interviews with former hostages: that in prolonged solitude, the social world makes its appearance 'as if it had always been on the inside' (p. 59), and so the illusory privacy of experience is countered for a time.

Ways of seeing easily congeal into conventions for seeing, and those conventions elicit agreement, so comfortable agreement may be a warning sign that closer examination is needed. The arts need to be part of the study of consciousness because they can both foster and expose psychological conventions. Art may help create our folk psychology: silently reading books may have contributed to our feeling that the 'first person' is a silent private voice in the head; film voiceovers (where we hear the character's voice as if we were listening to their thoughts) may have encouraged us to believe that thoughts are spoken in our heads. But many kinds of art are also good at 'sabotaging the familiar' (p. 185), and so may help in the fantastically difficult task of separating the 'how it seems' from the 'how it is'.

For Saunders, this seems (at least some of the time) to be one of the main tasks for the study of consciousness – a task for which we need all the help we can get, whether from a careful rewatching of *Being John Malkovich* or from an interview about surviving prison by building an elaborate imaginary castle. When I am not only the inquirer but also the object of inquiry and the means of inquiry, there's an easy deadlock of strong intuitions incessantly reconfirming themselves. But a social perspective shifts the goalposts. Not only may my experience not have many of the qualities I tend to think it does, like interiority or unvocalised language, it may not even be mine at all.

Experience is not so easily divided up as yours or mine. My experience may also be your experience, particularly if we were together. All experiences have shared contours such that to call something *my* experience is only ever partially true. (p. 39)

If the mystery to be explained is not the origin of the personal, private redness-of-red but an intrinsically interpersonal phenomenon, the questions we need to be asking change too.

The consequences of this new position aren't fleshed out as clearly as they might be, however. What exactly does it mean for an experience to be only partially mine? Elsewhere Saunders insists on the unreplicable uniqueness of any given experience, and on the importance of being able to conceal from someone else (say, a prison guard) what I'm experiencing; here the privacy of experience doesn't seem illusory. Maybe, then, what is actually shared is not a specific experience but an amalgam of related experiences that together form a prototype: 'My captive experience would be unique to me, but how it is to experience captivity is part of the world about us and therefore common ground' (p. 207). But saying merely that 'what we call "experience" is common between us' (p. 207) is a much weaker claim than saying that no experience is ever wholly mine. The ambiguity makes it hard to tell what the way forward is: do the methods we tend to group under the heading 'introspective' need supplementing with others that emphasise the sociality of conscious experience, or are they unfit for purpose, even actively misleading, and so need replacing altogether?

Even more fundamentally, it sometimes seems that Saunders does not in fact believe the 'how it seems' and the 'how it is' can be prised apart. Part of the book's foundation is that objectivity is just one more subjective standpoint. Because subjective experience has no objective nature, you can't get at subjectivity from multiple points of view, for example by triangulating a mixture of different physiological, behavioural, and 'self-report' measures, as mainstream experimental psychology might try to. The quest for maximum objectivity is for Saunders a wild goose chase, but so, perhaps, is any attempt to move from appearance to reality: 'it does not make sense to ask what my experiences are *really* like as opposed to how they appear to me. Our descriptions of how it is for us are *real*' (p. 155). But if so, it's not clear what the point is of defamiliarisation through solitary confinement or art: my conscious experience is just what I take it to be, including in my distracted everyday life. It's a central dilemma for consciousness studies: we can proceed in one direction to 'first-person' and 'second-person' methods designed to find out the truth about conscious experience (as in neurophenomenology), or in another to heterophenomenology's agnosticism as to the reality of conscious experience, or in another to the illusionist thesis that consciousness only *seems* to exist. It remains a dilemma in this book.

Our fragmented disciplinary landscape makes a lot of things more difficult. One of its unfortunate consequences is that people who argue for more integration often end up in a paradoxically divisive kind of combatism. For example, calls for more sciences-arts-humanities dialogue often sideline the 'mainstream' science that's been getting all the attention (I've been guilty of this in the other direction: thinking cognitive science can singlehandedly save literary studies.) Because Saunders starts by calling for us to expand our definition of the material, he ends up ignoring the material almost entirely. This is

unfortunate not just because the traditionally material – bodies and suchlike – is obviously crucial to any understanding of anything about being human, but also because fields that have been arguing for a different expansion of the material, e.g. from brains to the rest of bodies, have for a while now been tackling some of the trickiest problems and exploring some of the most interesting avenues that a social account also encounters. For example, Saunders devotes quite a bit of attention to the idea that drawing distinctions between things for the purposes of analysis doesn't mean that divisions exist in the real world. If only we could think of body and mind as we've learned to about nature and nurture, i.e. as interacting parts of the same system, we might find it easier to think about consciousness as we do about, say, identity: as obviously both physical and developmental. But although divisions like mind-body, self-other, me-world, etc. are clearly porous, it may still make sense to analyse them as systems with separable subparts that can operate with varying levels of independence. This is a crucial question in '4E' or distributed cognition, as is the question of exactly how to move away from neural representation as cognitive explanans. Dialogue with the scientific study of enactive and extended cognition would also benefit the investigation how we actively create experience rather than passively having it happen to us, or how a given experience depends on knowledge of other potential experiences, or how we think through the act of writing. Periodically declaring the irrelevance of the physical (for example in understanding the difference between freedom and captivity) certainly goes against the effort to bring all kinds of former backgrounds and foregrounds into equal focus.

And this selectivity goes right down to the experiences themselves. Saunders suggests that the kinds of conscious experience we choose to study may determine which theories work, or are useful: if you focus on sensory experiences like smelling a wild flower, you might well come to quite different conclusions about what's needed for a theory of consciousness than if you study how it is to feel fear. Fear, Saunders says, 'will be more useful to this book' (p. 129), because it foregrounds the fact that experience is a property of a person who stands in a relation to other persons. But that does make one wonder whether the appropriate approach wouldn't rather be to take a wide range of experiences and try to work out a theory that does justice to all of them, rather than just being 'partial' in a different way from usual.

It's really hard to do science-humanities interdisciplinarity well, for many reasons, one of which is that one often feels compelled to spend so long justifying it that there's not enough time left for doing it. Saunders rightly points out that much of the current discussion in consciousness studies stops right at the point where it most needs to go further. Possibly because my background is in literary studies, this is what I was left feeling about the examples from the arts given here. They're fascinating, as far as they go. Terry Waite says that after his release he despaired of being able to capture in words his experience of captivity until he realised that John Fowles's *The French Lieutenant's Woman* offered the only conceivable structure. Saunders observes that *The Wire* offers a glimpse of the polyphonic reality of consciousness which is so easily overlooked. He insists on the importance of a historical perspective that shows our preconceptions to be transitory. He links the double vision of consciousness (being conscious of something in some way depends on knowing you're not conscious of other things in other ways) with the combination of immersion and reflection in aesthetic engagement and creation. But advocacy and illustration need to be followed by close analysis and precise conclusions. So to engage with the *arts*, we should draw on the *humanities* (the disciplined study of the arts). The cognitive humanities have already started to bridge the gap by asking how textual structures interact with cognitive responses, and so how relationships emerge between minds in texts, minds in readers/viewers, and the folk psychology through which readers and viewers conceive of their own minds. The same goes for language, which gets lots of attention here: comparing the different verbs we apply to experience, distancing oneself from standard visual metaphors for experience, are great first steps, but the next would be to see how cognitive linguistics could illuminate the dynamics of those reciprocal constraints.

Maybe, like every book reviewer, I'm going to have to conclude that 1) the book should have been twice as long and 2) it should have been the book I would have written. But no matter. This is one you should read, especially if you don't think it's your kind of thing. It will give you a rewrite of the hard problem in paint, and it will make you think, and feel, about life, and yourself, and how easy it is to take freedom, and consciousness, for granted unless you make an active effort to suspend belief.

Emily T. Troscianko, 2017