Consciousness as depiction

The World in My Mind, My Mind in the World: Key Mechanisms of Consciousness in People, Animals and Machines by Igor Aleksander, Imprint Academic, 2005. £17.95 (UK)/$34.90 (US) (196 pp.) ISBN 1 845 40021 6

Ron Chrisley

Centre for Research in Cognitive Science (COGS) and Department of Informatics, University of Sussex, Falmer BN1 9QH, UK

Corresponding author: Chrisley, R. (email: ronc@sussex.ac.uk).

It has become a commonplace for reviews of books on consciousness to ask: Do we really need another book on consciousness? But in reviewing Igor Aleksander's latest book, I can brush this question aside, as there are relatively few books on its particular sub-topic: machine consciousness (other notable exceptions include, e.g. [1,2]). 'Machine consciousness' is perhaps a misleading term – it conjures up clunky robots from 1950s B-movies. The term 'artificial consciousness' might do just as well if not better, and this is particularly so in discussing Aleksander's work. Not only is it concerned with the engineering task of creating artificial systems that either model experience or are themselves truly aware; it is also concerned with the scientific task of illuminating natural consciousness by means of such work, a synthetic component that is often lacking in books on consciousness.

Although Aleksander has been investigating related issues for many years, his work is part of a recent resurgence of interest in machine consciousness, evidence for which can be found in the steady flow meetings on the topic: Cold Spring Harbor (2001), Skövde, Sweden (2001), Memphis (2002), Birmingham, UK (2003), Turin (2003), Antwerp (2004), Hertfordshire (2005) and Lesvos, Greece (2006). More specifically, Aleksander's work is an example of an approach that is currently finding increasing support in the machine consciousness community: the imagination or simulation approach (e.g. [3,4]). On this view, a key aspect of consciousness is the ability of an agent, say a robot, to 'imagine' (represent in a sensory-motor-grounded manner) experiences it is not currently having (e.g. the ability to answer the question 'what would I see if I turned my camera this way?'). The robot can then use such expectations, in conjunction with some basic affective system ('bumper sensors being activated is bad') to assist it in deciding what to do next ('I shouldn't reverse, because if I did, my bumper sensors would be activated'). Aleksander's book is not a summary or analysis of the different work being done in this sub-sub-field, nor was it intended to be. But it does set out to do three things that need doing: (1) clarify the approach, giving specific mechanisms when possible; (2) show how the approach can be used to model or explain various features of consciousness; (3) defend the approach against some standard philosophical objections.

A specific form of the first problem that all workers in this field must confront is: What makes this approach different from familiar AI models of mind? Most if not all traditional AI systems are capable of representing hypothetical situations and reason about them in a way that affects action (planning). Clearly the qualifier that the representations involved be somehow *grounded* in sensory-motor activity is meant to play a fundamental role in the novelty of the imagination-based approach. But that is more a slogan than an actual answer. Fortunately, Aleksander has more to say on the issue. He offers five 'axioms' that define what he takes to be essential for consciousness. They are axioms in the sense that they are not to be proven, but taken as an (introspectively plausible) stating point for enquiry, and are ultimately to be judged by the success or failure of the explanatory enterprise for which they form the foundation. Central to these axioms, and the book itself, is the notion of depiction: The representation of a world as being 'out-there', spatially and ontologically distinct from the subject. The other axioms concern the ability to manipulate depictions of, e.g. previously experienced situations, do so in an attentionally selective way, and use such depictions, in conjunction with an emotive or affective component, to determine action. So how can one create an artificial system that uses representations that are genuinely depictive? Aleksander prefers to explain by example at this point, but it seems that the essence of depiction is that items are represented in a way that is sensitive to the spatial implications of one's own actions. Thus, a representation of a fly when looking to the right will be different from the representation of an identical fly seen on the left, because the representation will reflect not only the qualitative properties of the fly, but also the different muscular signals indicating the different eye positions. So activating one of these representations as opposed to another will not just make a fly available in experience, but 'a fly that I have to move my eyes this way, rather than that way, to see'. Aleksander cites the discovery of 'locking' cells (e.g. gaze locking in monkey visual cortex) as evidence that something like his mechanism may underlie depiction in living systems. Those of an engineering bent may find the descriptions of depiction and other mechanisms frustratingly vague, perhaps even after reading the technical details that have been relegated to a nine-page appendix. But such readers can refer to Aleksander's technical papers; most will instead see the lack of technical detail as a welcome attempt at maintaining intelligibility.

The heart of the book is spent addressing the second task. Aleksander applies his axiomatic/depictive approach to explain aspects of sleep and dreams, confront the question of animal minds, resolve the dispute over whether consciousness is an illusion, and find a place for

the will in the face of Libetarian [5] (definitely not *libertarian*) and Wegnerian [6] scepticism.

Towards the end of the book, Aleksander attempts to anticipate and refute philosophical objections (the third task), taking the challenges offered by David Chalmers and Raymond Tallis to be his main obstacles. Although there is no space here to do justice to Aleksander's efforts on this point, I expect that although materialists will find themselves nodding their heads while reading these sections, few of the people who believe that consciousness requires a dualist approach will be convinced otherwise.

Although the emphasis is on machines, robots, and mechanisms, Aleksander would not call his approach 'computational', because for him that refers exclusively to symbolic, rule-based AI approaches; his own approach is based in sub-symbolic neural information processing. Furthermore, his book makes a welcome change from the (admittedly valuable) cataloguing of neural phenomena and their experiential correlates. It demonstrates that those who advocate the possibility (or existence!) of machine consciousness need not be behaviourists; one need not appeal to a consciousness variant of the Turing test to claim that an artificial system is conscious. Rather, one can examine the mechanisms underlying any conscious-like behaviour: if they resemble the mechanisms implemented by our neurophysiology, with a general structure as revealed in introspection, then one can truly deem it a conscious machine.

References

1 Franklin, S. (1995) *Artificial Minds*, MIT Press

2 Holland, O., ed (2003) *Machine Consciousness*, Imprint Academic

3 Cotterill, R.M.J. (2003) CyberChild: A Simulation Test-Bed for Consciousness Studies. *J. Consc. Stud* 10, 31–45

4 Hesslow, G. (2002) Conscious thought as simulation of behaviour and perception. *Trends Cogn. Sci.* 6, 242–247

5 Libet, B. *et al.* (1983) Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain* 106, 623–642

6 Wegner, D.M. (2002) *The Illusion of Conscious Will*, MIT Press