

Review

Home > Reviews >

Mind and Mechanism

McDermott, Drew V., *Mind and Mechanism*, The MIT Press, 2001, 262 pp, \$32.95 (hbk), ISBN 0-262-13392-X.

Reviewed by Varol Akman, Bilkent University, Ankara

2002.05.04

Share:

When I finished reading this book, I was reminded of the opening stanza of *Spirit Song Over the Waters*, a poem by Johann Wolfgang Goethe:

The soul of man resembleth water: From heaven it cometh, To heaven it soareth, And then again To earth descendeth, Changing ever.

I now think that McDermott's closing lines (p. 241)—and the general tone of the last chapter (Chapter 6)—were instrumental in elevating my spirits:

We must look for a coherent synthesis of our religious intuitions and scientific discoveries, so that as the very definition of what it is to be human changes, we will have a hope that our moral intuitions will not degenerate; or if not a hope, at least a prayer.

I'll later revisit this inspiring chapter and say a few more things about it. But let me start off this review with the usual question that faces any book reviewer. What is *Mind and Mechanism* about? Despite the experience I have just related, this book is not about ethical or religious affairs. (As a matter of fact, these arise only in Chapter 6.) A well-known AI researcher, McDermott is interested in computational approaches to consciousness. His reason for working in the field of AI is to solve the mind-body problem. In other words, his long-range goal is to understand how a physical organ, viz. the brain, can have experiences. This is a tough project because it involves the elucidation of the relationship between our mentality and the physical foundation of our body. How can a biological/chemical system (i.e., the human body) have beliefs, desires, intentions, and so on? Physicists have persuasive reasons to make us believe that ours is a material world (of particles at the bottommost level) that obeys physical laws. Once we commit ourselves to this view, it sounds quite bewildering—even enigmatic—to think that there is a place for independently existing minds in such a world.

When physicists speak, McDermott listens. As he makes clear on the first page of Chapter 1, his hypothesis is that we are all contraptions designed by evolution. The concept of mind arises because people's brains are biological computers. Crudely put, minds are produced by computers. The whole book thus becomes an extended argument about how one can be more specific about the way this production is realized.

The book consists of the following chapters: 1.The Problem of Phenomenal Consciousness (27 pp.) 2.Artificial Intelligence (63 pp.) 3.A Computational Theory of Consciousness (44 pp.) 4.Objections and Replies (29 pp.) 5.Symbols and Semantics (48 pp.) 6.Consequences (27 pp.)

Chapter 2, the longest, is a masterly account of the state of the art in AI. Discussed in this chapter are computational mechanisms for such classical problems of AI as game-playing (computer chess being the classic example), neural networks (including a clear discussion of the tensions between the symbolic vs. nonsymbolic approaches to AI), computer vision, robotics, speech-recognition/language-understanding, and automated theorem-proving. At first glance, this chapter looks like a stranger among the others. After all, this is the only chapter that reports technological advances (along with their underlying theoretical foundations, to be fair). But the author has a good reason for including it: to draw a clear boundary between what we can now build and what is sheer speculation (science fiction). Thus, when McDermott deliberates about computational mechanism in various other places in the book, he in fact is referring to mechanisms like the ones mentioned in Chapter 2.

McDermott sees mind as a self-fulfilling description, a description that brings mind into being. He offers a fine analogy by considering the windows in the user interface of a personal computer. These windows are there because the computer is working in a way supporting their existence. It is able to do so by running procedures that use data structures—textual descriptions portraying what the windows are to be like. In a nutshell, a formal description starts to act causatively when interpreted by a computer. In some sense, the mind is also a window, albeit one with a Herculean description.

McDermott assigns great importance to models of the world. These models include models of the self. The self-model is the source of everything one knows about oneself. More interestingly, it is because of the self-model that one believes in free will. Subtract the self-model and you would end up with an entity which is not a fully functioning person anymore.

For McDermott, the problem of phenomenal consciousness—how a physical object can have experiences—is the toughest nut to crack. To put the matter differently, explaining what it is to have qualia is the hardest problem. To this end, McDermott first tries to demystify the notion of free will. His insight is that a robot must model itself in a different way from other objects in its world in order to avoid infinite regress. Consider a straightforward cycle of events in the life of a robot, starting with perception, leading to making a tentative prediction, and concluding with revising an action. McDermott claims that this chain of events cannot be accurately embedded in a model unless the symbols denoting the robot itself are flagged as free from causality. (You'll have to read Chapter 3, pp. 97-98, to appreciate this fully.)

This approach to explaining free will can also be used to explain away qualia. The key move is to notice that a sequence of goals must come to an end with a goal that can't be further questioned. Imagine a robot on a dangerous mission to save people from drowning in a river. As long as the robot is immersed in water, its underlying system can label wetness as “undesirable but OK for the time being.” According to McDermott, at this point the robot's apprehension of wetness is analogous to a quale of disagreeableness. In representing this state, the robot classifies it as “to be shunned or evaded as much as possible.” To cite another of McDermott's examples: “[A] robot may dislike going into burning buildings because it dislikes heat. But it doesn't dislike heat because of further bad consequences; high heat is *intrinsically* not-likable” (p. 102, my emphasis).

But then there is no need for qualia in the computational system of a robot. Consciousness arises through the employment of a self-model and qualia are occasioned by the process of self-modeling. That means that what is manifested by our robot is virtual consciousness, which eventually boils down to physical events.

Chapter 4 is devoted to a thorough defense of these proposals (formulated in Chapters 1 and 3) against various (some even classical) objections. Discussed in Chapter 5 are numerous full-fledged or partial theories of consciousness or related matters by influential philosophers, e.g. (in no particular order), Tye, Carruthers, Clark, Rey, Block, Chalmers, Jackson, Dennett, Shoemaker, Nagel, Lycan, Searle. Almost all of McDermott's arguments and analyses in this chapter are to the point, upright, and illuminating. His grasp of the contributions of the aforementioned people is commendable in its clarity.

Chapter 5 is somewhat technical and treats a question which is not central to McDermott's general theory: the observer-relativity of symbols and semantics. A notable proponent of observer-relativity, Searle famously said, “For any program and any sufficiently complex object, there is some description of the object under which it is implementing the program. Thus for example the wall behind my back is right now implementing the Wordstar program, because there is some pattern of molecule movements that is isomorphic with the formal structure of Wordstar.” McDermott's refutation of this dangerously relativistic viewpoint is based on his painstaking analysis of the notion of decoding, a mapping from a computer's states to its computational realm. It is then argued that while everything might be considered as a computer sometimes (with respect to some decoding), this fact does not endanger the concept of a computer. Using a clever continuity argument (small state perturbations causing little difference in the output), McDermott shows that if someone makes a Searlean claim of the sort just mentioned, then the burden of proof is on him to demonstrate that the continuity requirement is not violated.

Assuming that McDermott's general views regarding the self and qualia are correct, an infamous question raises its head: is there anything precious or sacred about a person (= a machine)? The history of AI is replete with numerous instances of this question because people are (understandably) worried about the moral dimension of the AI enterprise. The question that has been asked for the umpteenth time, most recently in the flashy Spielberg movie *A.I.*, is this: can an artificial intelligence “be the subject or object of moral judgment?” (p. 217). McDermott states that a robot can have phenomenal consciousness but lack morals. If I understand him correctly, his argument for this is based on the behavior of a program. Assume we build a robot which is just like an average person when it comes to attributes such as love, morality, aesthetics, humor, etc. Now, it is conceivable that by just tweaking some program variables, we could get radically different (perhaps unrecognizable) versions of these attributes. Consequently,

We'll take robot aesthetics [similarly, robot morals -VA] seriously if the robots ever reach a point where we can debate them, destroy them, or lobotomize them, but can't change their tastes [moral codes -VA] by adjusting a few lines of code. (p. 221)

The last big question that McDermott grapples with is belief in God in the absence of dualist nonphysical realms. His stance is not anti-religionist; he doesn't think that religion is a menacing force. Neither does he consider belief in God as a potentially short-lived activity in the broad development of civilization. McDermott is convinced that mankind will always be troubled by finiteness and long for the comfort provided by the unfathomable or incomprehensible. In a memorable passage, he writes:

The only way to reconcile God's silence [remember Wittgenstein's closing remark in Zettel: “You can't hear God speak to someone else, you can hear him only if you are being addressed.” -VA] with his existence is to assume that he poured himself into the world when he created it. His intervention in the world consists in creating laws of physics that allow complex physical systems to evolve that understand those very laws; in making it possible for weird creatures to evolve from monkeys and grasp the need for an overarching set of moral principles. [pp. 237-8, my emphasis]

McDermott's God is presumably one who does not answer prayers or does not prevent bad things from happening to good people. It is a God who turned “us loose in such an uncaring world” (p. 239). But he takes a reverent stance: the only way for us is to accept the world as given and thank God for our existence. Thinking about what God wants us to do is good because otherwise there is a real danger of nonchalance and aberrant behavior (p. 238).

To wrap up, one strength of this book is its almost encyclopedic coverage; it simply looks like the agenda of a busy philosopher of mind-language-and-morals. And while it may be surmised that in treating so many diverse concepts it is easy to fall into the trap of superficiality, I did not really notice anything of the sort. McDermott has this natural gift for explaining the knotty or perplexing with such grace and vividness that one cannot help but admire his virtuosity. Thus, his modesty and worries about not “using the usual philosophical tools to approach [philosophical questions]” (p. 24) or not conducting his discussions “in the pure philosophical style” (p. 25) are definitely misplaced.

Do I see any shortcomings of *Mind and Mechanism*? Well, probably just this one. Except for a passing remark (on something else), McDermott does not connect his proposal with the doctrines of Davidson. This is a pity because such a comparison would be promising, even fruitful. As many readers will know, Davidson has worked out an ingenious answer to the puzzle of the mental. Basically, he takes it for granted that the essential properties of matter as described by physicists are the only properties we have. Thus, he subscribes to some form of materialism. However, he thinks that one can be a materialist while also asserting that the mental cannot be reduced to the physical. Assume that you have complete knowledge in front of you of your brain and any relevant neurophysiological systems. According to Davidson, this knowledge cannot constitute knowledge of your beliefs, desires, intentions, etc. This he maintains without really taking a dualist stance, that is, without assuming that your mind has a separate kind of existence. Rather, his point is that our vocabulary for describing the mental does not match the concepts of physics in the right way. For example, he sees the principle of rationality as a most crucial aspect of the mental (especially belief), and holds that this principle has no echo in physical theory. Davidson's thesis is that the nature of mental phenomena does not permit law-like regularities connecting the mental phenomena with physical events in the brain.

In the preface of his *The Seas of Language*, Michael Dummett offers useful advice. He hopes that readers might enjoy his papers and be stimulated to new approaches of their own. More importantly, he says that McDermott does not expect agreement. I think the same comments can be made for *Mind and Mechanism*. In the space of a merely 250 pages McDermott spans the landscape of almost every important problem in computation, cognition and mind, while not causing indigestion (but presumably a lot of disagreement). A capable author in lucidly explaining the most intriguing phenomena, he formulates fine—but tentative—solutions for some of these vexed questions. If you want an intuitive (and frequently deeper) grasp of what is keeping philosophers of mind busy nowadays read this book. With the recent flurry of books and conferences (see <http://www.wired.com/news/technology/0,1282,51765,00.html> for a very recent one) on consciousness, it seems we are closer to filling the gap that discouraged William James. “We know what consciousness is,” he once wrote, “as long as no one asks us to define it.”

Share:

Search Reviews Keywords Search