

## Book Reviews

Hans Moravec, *Robot. Mere Machine to Transcendent Mind*, New York, NY: Oxford University Press, Inc., 1999, ix + 227 pp., \$25.00 (cloth), ISBN 0-19-511630-5.

In a follow-up to his 1988 book *Mind Children: The Future of Robot and Human Intelligence*, Hans Moravec tells the tale of how the robots shall inherit the Earth, and the rest of the universe as well. Such stories are common in science fiction, the difference here is that the story is told by a leading researcher in robotics – Moravec was a founder of the Robotics Institute at Carnegie Mellon University.

As in his first book, Moravec's twist is to embrace this seemingly scary future. This is because he sees the robots as the next step in our cultural and evolutionary development. As the book slides from history into prophesy, it is laced with interpretations of the stranger results from theoretical and experimental physics, and ends with a sketch of Moravec's vision of the world spirit. The book itself is neither science nor fiction, but something more like a theology for techno-optimism. And if you thought Jean Baudrillard's (1983) *Simulations* was an over-stated caricature of techno-capitalism's late 20th century world-view, prepare yourself for the flag-waving account of a zealot.

Moravec starts out Chapter 1, "Escape Velocity", with a sort of Myth of Origins, the story of humanity's first steps down the path of civilized culture after the last ice age. It's the familiar story of how biological evolution endowed humans with large and flexible brains, and how language and cooperative living led to the division of labor and expertise, the accumulation of wealth and leisure, and eventually scientific thought. This is all meant to explain how civilization reached its current technological complexity.

Chapter 2, "Caution! Robot Vehicle!", gives a more recent history of the great strides taken by robotics research over the last 50 years, from Grey Walter's tortoises up to cars that can drive themselves across the country. In many ways, Moravec's personal recollections of his days at Stanford are the most interesting part of this book. But in telling this history he goes to great expense trying to apologize for the limited success of robotics by showing how slow, underpowered computers were to blame.

Chapter 3, "Power and Presence", sets up the rest of the book by presenting Moravec's argument for why computers will be so much faster in the not-so-distant future that robotics and AI will start displacing physical and intellectual laborers, including scientists. Projecting from the current rate of increasing computational power (using Moore's law), and making some rather sketchy estimations of the computational capacity of the human brain (based on the complexity of the retina



and its volume relative to the rest of the brain) Moravec predicts that CPUs as powerful as the human brain will be on desktops by 2040.

And not only will the computers be as powerful as human brains, Chapter 4, “Universal Robots”, explains how AI researchers will make them smarter than humans too. As computational power approaches brain-like proportions, AI will effectively recapitulate the cognitive evolution of primate brains by solving the problems facing the construction of general purpose robots. First will come basic perceptual and motor skills, then effective reasoning skills, then basic concept learning, then complex learning and abstract metaphorical reasoning. Robots will become increasingly viable commercially – first for such applications as autonomous vacuum cleaners and later for all sorts of things humans dislike doing. As they are designed to do more sophisticated sorts of work, robots will develop towards a general purpose laborer, the *universal robot*, which will be able to learn new tasks and interact effectively with humans.

Readers of *Minds & Machines* may be disappointed to know that this ordering of future developments loosely follows a brain metaphor rather than current or expected progress in any areas of cognitive research. In fact, he qualifies his predictions of machine intelligence by saying that by 2040 we won’t be able to fully simulate the brain, only to run elegantly programmed AI algorithms. But apart from his discussion of robot vision, he offers no sense of how the different areas of AI research will achieve such successes – increasing computer power is the panacea for all the problems of machine intelligence.

Chapter 5, “The Age of Robots”, moves closer to the realms of science fiction. Moravec predicts that robot evolution will reach a critical threshold or “escape velocity” after which they will no longer be dependent on humans for their design, maintenance or manufacture – they will come of age. In short, robots will be maintaining civilization, and probably controlling it, as early as 2050. After they take over the physical and mental labor, they’ll quickly take over strategic decision-making in the corporations (a la *Neuromancer*). All of this shakes up the economic system a bit, as humans struggle to govern a world in which nobody has a job, the robots control the means of production, and consumer goods are obscenely plentiful – a utopian welfare state. Eventually the robots will get wise to their own exploitation and do something about it (as in *Russom’s Universal Robots*) and take control of everything unless we figure out how to stop them.

The smart new computers will also propel science ahead at warp speed, producing most of the technologies that *Star Trek* has been promising us. And of course, people will experiment by replacing body parts with robotics and their brains with computer chips in true cyberpunk fashion. A race of ex-human cyborgs will evolve, whom Moravec dubs “Exes”, but they will be denied human rights and sent into space (a la *Blade Runner*). The Exes will quickly colonize space and flourish there, bearers of the cultural destiny of humanity – our *Mind Children*. In fact, the first five chapters don’t say much that wasn’t in Moravec’s first book, he just tries to present

more scientific justification and philosophical argumentation for his predictions this time around.

Chapter 6, “The Age of Mind”, marks the departure from plausible predictions to pure prophesy. According to Moravec, the advancement of science will continue to accelerate, providing the means to organize the sub-atomic tangles which make up the space-time continuum into bits for computation and information storage. Compounding this with technologies for quantum computation, teleportation and time-tunneling logic-gates, there will evolve a kind of supermind, capable of permeating any material substance with its own mental computations. As such, these superminds will spread across the universe, transforming all available matter into extended bits of their minds in a desperate search for increased mental powers.

But even though the superminds may destroy all the Exes and the physical world as we know it, the revelations of Moravec’s final Chapter 7, “Mind Fire”, instruct us not to worry. Once taken up in the rapture of the mind fire, our descendants, and our ancestors for that matter, will live on eternally in these superminds as simulations and memories. The encoding of information will be so efficient that Earth and all its inhabitants can be converted into a life-like simulation and stored in a chunk of matter an infinitesimal fraction of the size of the original. In fact, the whole history of the world will be run and rerun in a virtual reality (VR) for purposes of research or entertainment for the superminds. And since these simulated realities will be as richly detailed as our reality, they will be just as real. So it might be the case that we are doomed to an eternal recurrence of our lives, or perhaps we will be free to choose alternate courses through the simulated worlds, or perhaps we will live out every possible life course. Whichever it is, the ultimate destiny of human culture will be a universal mind which eternally contemplates itself and all possible interpretations of subjective experience, a kind of VR-nirvana.

The readers of *Minds & Machines* will likely find many points of contention with Moravec’s fantasy, regardless of the discipline they come from. Computer scientists will wonder how computer power alone can solve all the problems in AI. Cognitive scientists, linguists and neuroscientists will wonder how vast amounts of computer power might help to solve any of the problems of the human mind. And philosophers will sense that the critical issues surrounding artificial minds have been merely gestured at.

A more generous reading of this book is as a survey of the philosophical issues surrounding artificial minds, and as an opportunity to discuss the central issues of our technological society: What is so desirable about a world in which there is no more work to be done? Is leisure more desirable than self-determination? Shall human culture always desire progress—bigger, better, faster? And why should the robots who take over the world desire ever increasing power, control and complexity? Is it always rational to pursue “progress” at any cost? Are simplicity and stability necessarily irrational?

If we agree that his technological predictions are even remotely correct, then

we are going to have to deal with important social and ethical issues regarding the status and rights of robots and AIs. While you get a sense of these issues from reading the book, Moravec makes little effort to start a thoughtful discussion of the matter. The displacement of workers by machines takes place with little struggle, and seems to lead uneventfully to a welfare state which appears to have no difficulty dividing up the plentiful goods produced by robot slaves. If such a transformation of labor were to occur, the economic and political ramifications would be profound, and would most certainly require more time than the decade that Moravec allows for it. Moreover, at one point he argues that we should keep the robots enslaved for as long as we can manage to, yet also makes the point that robots will be just as conscious and sensitive as humans. Are we meant to conclude that human and robot rights are merely a political convenience?

Moravec acknowledges that his speculations raise deep logical, metaphysical and epistemological issues, yet he only offers a page each on “Consciousness” and “Existence,” for instance. These are issues which the best philosophers have been wrestling with since Plato. Moravec also promotes an odd metaphysical position he calls “physical fundamentalism” which is elaborated as a way of being a realist about mathematical descriptions of physical reality. He claims at many points that there are no inconsistencies in his theory, only some counter-intuitive consequences. But alas, his theory seems to be nothing more than idealism, where the mind of God has been replaced by a VR simulation.

In short, Moravec maintains that every logically and mathematically consistent interpretation of the world is simultaneously true. Accordingly, our current experience of reality is just one such interpretation. A VR simulation could have an identical interpretation, as could a rock, with no essential difference between them. Whereas philosophers such as Hilary Putnam and John Searle have argued that third-person semantic interpretations are too promiscuous for identifying minds because they would allow rocks to be interpreted as having minds, Moravec happily asserts that rocks really do have minds and we simply choose not to interpret them in this way – consistent, but counter-intuitive. But can he really have his Platonism and be a physicalist too?

If he wants to be a materialist, then any interpretation of the world would require a materially-describable mental system with the semiotic capacities to generate such an interpretation, and the available interpretations would be limited by those capacities. But he clearly rejects this:

The difference between the physical and mathematical reality is an illusion of vantage point: the physical world is simply the particular abstract world that happens to contain us. ... The Platonic position on simulation answers that the abstract relationships that constitute the mind, including its own self-interpretation, exist independently, and a robot, a simulator, or a book describing the action, no less than the biological brain, is just a way of peeking at them. (Moravec, 1999: 196–7)

Which leaves him with a more or less straightforward idealism – only ideas are real. Interpretations are easy to come by since, like Plato's Forms, they are all floating eternally in heaven. And while they are interpretations *of* physical entities, they do not themselves require any material basis to exist. He admits, however, that not all possible interpretations are equal, some are *selected* while others are not. But who or what does the selecting? *Minds* of course, but this just begs the whole question of whether or not a simulated mind is a real mind, since minds are just another kind of interpretive pattern floating around in the void.

Even if Moravec's metaphysics could be fleshed out coherently, idealism is not without its own problems. What about epistemology, ethics and aesthetics? If there are only ideas, why do some ideas have different access to one another, and different relations to each other, to the truth and to conscious experience? Where does this leave the debate over human and robot rights? If one honestly believes that life is a VR-dream, then why should one struggle for the good or the beautiful? If all the possible-worlds are just different actual worlds, as Moravec suggests, then how could any event or decision ever matter since every alternative will be chosen in one world or another? And why bother building robots and AIs if rocks are already intelligent?

To his credit, Moravec has given philosophers a brand new set of thought experiments to play with. He also gives a clear presentation of many of the puzzling consequences of quantum computation. At least a dozen papers could be written to try to straighten out and criticize his new version of idealism and the techno-theistic cosmology which it implies. Also worthy of closer examination is the unintentionally stark image of the promised land of this cosmology. Intended to be utopian, Moravec's universe seems awkwardly dystopian. It's not just dehumanizing, it's dematerializing – the universe itself is just computation and simulation.

I would recommend the book to scientists and engineers looking for some entertaining reading, and to philosophers seeking a spokesperson for a rich set of positions to criticize. The book's greatest contribution is as an attempt to organize the metaphysics of science fiction into a coherent position with some scientific plausibility. As such, it should not be criticized too harshly for failing to be fully coherent or plausible. In short, it's the kind of book you love to hate.

PETER M. ASARO

*Beckman Institute*

*Departments of Philosophy and Computer Science*

*University of Illinois*

*Urbana, IL 61801, USA*

*E-mail: asaro@uiuc.edu*