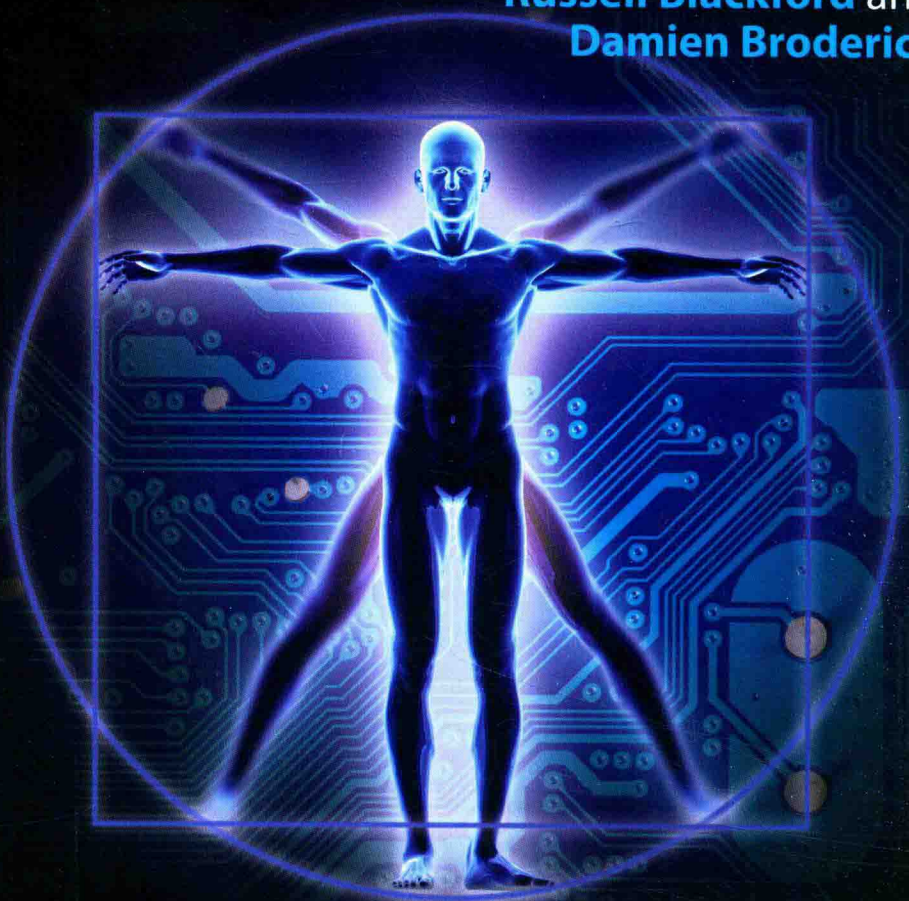
Intelligence Unbound

The Future of Uploaded and Machine Minds

Edited by

Russell Blackford and
Damien Broderick



WILEY Blackwell

Intelligence Unbound

The Future of Uploaded and Machine Minds

Edited by

Russell Blackford and
Damien Broderick

WILEY Blackwell

This edition first published 2014

Registered Office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Offices
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.

Library of Congress Cataloging-in-Publication Data applied for.

Hardback ISBN: 978-1-118-73641-8
Paperback ISBN: 978-1-118-73628-9

A catalogue record for this book is available from the British Library.

Cover image: Circuit board © sborisov /iStockphoto; Vitruvian Man © Devrimb /iStockphoto.

Set in 10/12pt Sabon by Laserwords Private Limited, Chennai, India
Printed and bound in Malaysia by Vivar Printing Sdn Bhd

1 2014

Intelligence Unbound

To Aubrey Townsend, who handed me the tools
Russell Blackford

To R. Daneel Olivaw, Golem XIV, Donovan's Brain, and Paul Durham, in the hope that things turn out better next time
Damien Broderick

Notes on Contributors

Nicholas Agar is a New Zealand philosopher, based at Victoria University of Wellington. His research is focused on ethical issues arising out of the application of new technologies to human beings. His most recent books are *Humanity's End: Why We Should Reject Radical Enhancement* (2010) and *Truly Human Enhancement: A Philosophical Defense of Limits* (2014).

Michael Anissimov is a futurist focused on such emerging technologies as nanotechnology, biotechnology, robotics, and artificial intelligence. He previously managed the Singularity Summit and worked as media director for the Machine Intelligence Research Institute, as well as co-founding Extreme Futurist Festival.

Stuart Armstrong and Seán ÓhÉigeartaigh work at the Future of Humanity Institute of Oxford University, where they analyze the major risks facing humanity, and how these can be prevented or mitigated. Recent work has focused on the risks and ethics of AI, human biases, and the reliability of predictions.

Russell Blackford is an Australian philosopher and literary critic. He is a Conjoint Lecturer at the University of Newcastle, NSW, and editor-in-chief of *The Journal of Evolution and Technology*. His recent books include *Freedom of Religion and the Secular State* (2012) and *Humanity Enhanced: Genetic Choice and the Challenge for Liberal Democracies* (2014).

James Bodington is a doctoral candidate in philosophy at the University of New Mexico, where he studies twentieth-century and contemporary continental philosophy, especially philosophy of religion and philosophy of

technology, with a particular emphasis on the ethical and political ramifications of technology.

Damien Broderick holds a PhD in the literary theory of the sciences and the arts from Deakin University, and has written or edited some 60 books in several disciplines, including a number of prize-winning novels. His *The Spike* (1997, 2001) was the first general treatment of the Singularity. In 2008 he edited an original science anthology *Year Million*, on the prospects of humankind in the remote future.

David J. Chalmers is Distinguished Professor of Philosophy and Director of the Centre for Consciousness at the Australian National University and Professor of Philosophy at New York University. He is best known for articulating what he has dubbed the "hard problem" of consciousness – explaining how physical brains and bodies give rise to "qualia," or subjective experiences. His best-known book is *The Conscious Mind* (1996).

Joseph Corabi is an Associate Professor of Philosophy at Saint Joseph's University. He has published numerous articles on philosophy of mind and philosophy of religion.

Linda MacDonald Glenn is a bioethicist, healthcare educator, lecturer, consultant, and attorney. She holds faculty appointments at the Alden March Bioethics Center and California State University Monterey Bay. Her research is focused on the sociopolitical implications of exponential technologies and evolving concepts of legal personhood.

Ben Goertzel, PhD, chief force behind the recent movement toward artificial general intelligence in the AI field, is chief scientist of financial prediction firm Aidyia Holdings and chairman of AI software company Novamente LLC and bioinformatics company Biomind LLC. His research work encompasses artificial general intelligence, natural-language processing, cognitive science, data mining, machine learning, computational finance, bioinformatics, virtual worlds, gaming, and other areas.

Kathleen Ann Goonan is the author of *Queen City Jazz* (1994), *The Bones of Time* (1996), *Mississippi Blues* (1998), *Crescent City Rhapsody* (2000), *Light Music* (2002), *In War Times* (2007), *This Shared Dream* (2011), and *Angels and You Dogs* (2012). She is a Professor of the Practice at Georgia Institute of Technology, Atlanta, where she teaches creative writing and examines the intersection of culture, science, technology, and literature. Her website is www.goonan.com.

Victor Grech is Consultant Pediatrician (Cardiology), Pediatric Department, Mater Dei Hospital, Tal-Qroqq, Malta, and author of several searching essays on the thematics of *Star Trek*.

Robin Hanson is an Associate Professor of Economics at George Mason University and a research associate at the Future of Humanity Institute of Oxford University. He is known as an expert on prediction markets and was a principal architect of the Foresight Exchange, DARPA's FutureMAP project, and IARPA's DAGGRE project.

James J. Hughes is the Executive Director of the Institute for Ethics and Emerging Technologies, and a bioethicist and sociologist at Trinity College in Hartford, Connecticut, where he teaches health policy. Hughes is author of *Citizen Cyborg*, and is working on a second book tentatively titled *Cyborg Buddha*.

Randal A. Koene introduced the multidisciplinary field of whole brain emulation and is lead curator of its scientific roadmap. He is founder of the Carboncopies.org foundation and neural interfaces company NeuraLink Co, and Science Director of the 2045 Initiative. His publications, presentations and interviews are available at http://randalkoene.com.

Richard Loosemore is a lecturer in the Department of Mathematical and Physical Sciences at Wells College. He graduated from University College London as a physicist and from Warwick University as a cognitive scientist. His background includes work in artificial intelligence, cognitive science, physics, software development, philosophy, parapsychology, and archeology.

Max More, who received his PhD in philosophy from the University of Southern California, is a strategic philosopher recognized for his thinking on the implications of emerging technologies. More's contributions include founding the philosophy of transhumanism, developing the Proactionary Principle, and co-founding Extropy Institute. He is currently president and CEO of the Alcor Life Extension Foundation.

Seán ÓhÉigeartaigh and Stuart Armstrong work at the Future of Humanity Institute of Oxford University, where they analyze the major risks facing humanity, and how these can be prevented or mitigated. Recent work has focused on the risks and ethics of AI, human biases, and the reliability of predictions.

Nicole Olson is a Canadian transhumanist writer/researcher holding a bachelor's degree from the University of Alberta in philosophy and sociology.

Massimo Pigliucci is a Professor of Philosophy at the City University of New York. His research is concerned with philosophy of science, the relationship between science and philosophy, and the nature of pseudoscience. His publications include several books, most recently

Answers for Aristotle (2012) and *Philosophy of Pseudoscience* (co-edited with Maarten Boudry, 2013).

Joel Pitt, PhD, is a scientist and software developer based in Wellington, New Zealand. As a scientist, he has contributed original research to molecular biology, machine learning, and ecology. As a developer he has been the CTO Demand Analytics, and currently works for Dragonfly Data Science, a science consultancy in Wellington.

Anders Sandberg has a background in computational neuroscience and the ethics of human enhancement. Since 2008 he has been James Martin Research Fellow at the Future of Humanity Institute at Oxford University, where he is investigating neuroethics, global catastrophic risks, and applied epistemology.

Susan Schneider is an Associate Professor of Philosophy at the University of Connecticut. She has published many articles in the fields of metaphysics and philosophy of mind as well as *The Language of Thought, The Blackwell Companion to Consciousness* (with Max Velmans), and *Science Fiction and Philosophy*.

Joe Strout's career blends science and technology, with degrees in psychology and neuroscience, and extensive experience as a software engineer. He works now as a software consultant, developing artificial intelligence algorithms for the game industry, as well as other applications in business and medicine. His website is http://www.ibiblio.org/jstrout/uploading.

Iain Thomson is Professor of Philosophy at the University of New Mexico. The author of two books, *Heidegger on Ontotheology: Technology and the Politics of Education* (2005) and *Heidegger, Art, and Postmodernity* (2011), Thomson has published dozens of articles in philosophical journals, essay collections, and reference works, and his writing has been translated into seven languages.

Natasha Vita-More, PhD, is a Professor at the University of Advancing Technology and Founder of H+ Lab. She has appeared in over 24 televised documentaries and featured in *Wired*, the *New York Times*, and *Village Voice*. She is chair of Humanity+ and a Fellow of the Institute for Ethics and Emerging Technologies.

Mark Walker is an Associate Professor in the Department of Philosophy at New Mexico State University, where he holds the Richard L. Hedden Chair of Advanced Philosophical Studies. His book, *Happy-People-Pills for All* (2013) argues for creating advanced pharmaceuticals to boost the happiness of the general population.

Naomi Wellington is a postgraduate philosophy student at the Australian National University, working under the supervision of Daniel Stoljar and David Chalmers. Her academic background includes a BA with philosophy honors (H1) from Monash University. Her primary areas of interest are philosophy of mind and philosophy of neuroscience.

Contents

Introduction I: Machines of Loving Grace (Let's Hope)

Damien Broderick

1 Machine minds or humans copied into machines?

In an immensely confident but typical summary of the neurocomputational model of mind now dominant in science, Nobel Laureate Eric Kandel wrote in 2013:

> This new science of mind is based on the principle that our mind and our brain are inseparable. The brain is a complex biological organ possessing immense computational capability: it constructs our sensory experience, regulates our thoughts and emotions, and controls our actions. It is responsible not only for relatively simple motor behaviors like running and eating, but also for complex acts that we consider quintessentially human, like thinking, speaking and creating works of art. Looked at from this perspective, our mind is a set of operations carried out by our brain.[1]

More than two decades earlier, the science fiction writer Charles Platt offered a somewhat ampler view:

> A person's mind is structure as well as content. Without the structure, the content can't function. Our minds have to have the specialized architecture ... in which to operate. We can store our brain data elsewhere, but when we do that, it's as nonfunctional as a videodisc without a disc player. (Platt 1991: 238)

In the next 25 to 100 years, genuinely intelligent machines are likely to be developed up to and beyond the highest levels of human ability.

Intelligence Unbound: The Future of Uploaded and Machine Minds, First Edition.
Edited by Russell Blackford and Damien Broderick.

We're not there yet, in part because the raw computational power of the brain hugely outstrips even the fastest computer. In mid-2013, the world's top supercomputer was the Tianhe-2, holding more than a million gigabytes of memory and running at some 50 petaflops (where a petaflop is a thousand trillion calculations per second) on its best days. Using an only slightly less extraordinary machine, Japan's 10 petaflop K supercomputer, scientists simulated 1 percent of 1 second of human brain activity. That took 40 minutes of screamingly fast calculations.[2]

You would need to multiply that by a factor of a quarter million to emulate a brain. Luckily, Moore's law (roughly: "computer power doubles every year and a half"[3]) suggests that a machine of this majestic status will be available – all things going well – in perhaps 30 more years. And of course in the meantime, scientists might learn better ways to get the job done sooner on leaner computers.

> Markus Diesmann of the Institute of Neuroscience and Medicine at Germany's Forschungszentrum Julich believes that, within the next decade, we'll be able to use exascale computers – capable of 1000 times one quadrillion operations per second – to represent the entire [*sic*] of the brain "at the level of the individual nerve cell and its synapses."[4]

And Henry Markram, a professor of neuroscience at the Swiss Federal Institute for Technology and founder and director of the Blue Brain Project, is coordinating the Human Brain Project (Keats 2013). This 10-year, €1.3 billion flagship project, selected in January 2013 by the European Commission, plans to simulate

> a rat cortical column. This neuronal network, the size of a pinhead, recurs repeatedly in the cortex. A rat's brain has about 100,000 columns of [about] 10,000 neurons each. [A] human cortex may have as many as two million columns, each having [about] 100,000 neurons each ... These models will be basic building blocks for larger scale models leading towards a complete virtual brain.[5]

Will they necessarily be conscious, such brainy machines? Perhaps not, or at any rate not as we experience consciousness. The speeding locomotive, or "Iron Horse," never resembled a real horse, yet it carried a heavier load and moved much more swiftly and without tiring. A submarine isn't much like a whale, yet dives deeper and travels faster. Birds sing more beautifully than jet planes or rockets, but their capacity to fly high, far, and rapidly was outstripped by machines a century ago. Chess programs defeat grand masters without being self-aware, and IBM's Watson supercomputer beat top human contestants on *Jeopardy!*, winning a million dollars but without a jitter of anxiety or a shout of joyful pride.

Even so, machine or artificial intelligence (AI), unlike ours, might well have the ability to understand, modify, and improve its own source code, carrying it by great leaps into domains of ability that unaided flesh can never hope to reach. Half a century ago, the mathematician I.J. Good proposed that an "ultraintelligent machine" could design ever more enhanced versions of itself, resulting in an "intelligence explosion" that would leave humans far behind (Good 1965). If such supersmart computers also achieve consciousness, we (or our children and grandchildren) shall share the planet with a new and intriguing species of mentality.

But wait – what *is* intelligence? Thousands of learned books and scientific or philosophical papers have probed every corner of this apparently simple question with no clear consensus emerging. We can start with theoretical neurophysiologist William Calvin's breezy summary in *How Brains Think*:

> I think of intelligence as the high-end scenery of neuro-physiology – the outcome of many aspects of an individual's brain organization which bear on doing something one has never done before ... some of *what* intelligence encompasses are cleverness, foresight, speed, creativity, and how many things you can juggle at once. (1997: 11)

Instead of our brutally slow chemical neurotransmitters, ionic currents, and neural designs, built by millions of years of ad hoc evolution, AI will use engineered electronic or photonic neural nets operating a million times faster. Instead of memories limited by the gene-architected size of our skulls and the human birth canal, AIs will possess effectively limitless storage constrained only by pathways traversed at the speed of light. In that sense, the arrival of advanced AIs will mark the end of some of the limitations that bind human intelligence. Intelligent and superintelligent machines will truly represent "intelligence unbound."

If and when this happens, humanity will face ethical issues of unprecedented gravity and difficulty. What obligations do we owe to artificial minds? Can they morally be switched off, like any other instrument or mechanical device? Or do they share human rights to life and the pursuit of happiness, the right of due process? Is there any way in which their designers can defang hazardous AIs that might turn on us, can make them compliant, obedient to their creators? Or is that slavery, mind bondage? If they are our intellectual superiors, can they at least be encouraged to adopt an attitude of benevolence toward us? Is it even technically possible to enforce friendship between protein and silicon beings, once the AIs pass beyond human comprehension in their abilities and potential?

In addition to this vexed and giddy outlook, in a near future of such fabulous machines, it will be possible to blend human and machine by enhancing

our current bodies with chips, modules, and interface devices (a process of "cyborgization" that has already begun).

All these prospects, and more, are discussed in detail in the chapters of this book. No single viewpoint is privileged throughout; these topics remain genuinely controversial, even philosophically troubling, so it is necessary to approach the topics carefully, exploring the pros and cons. And if the imminent arrival of machines with intelligence, however alien, is sure to throw our world into confusion and tumult, how much more will the possibility of minds copied from organic brains to inorganic machines? Not just copied as a static representation, as the Mona Lisa might be counterfeited with great fidelity by a skilled artist, but imbued with emotion, awareness, and all the other aspects of personhood.

2 Emulating the mind

This radical option might become available alongside the emergence of machines powerful enough and intricately connected enough to house a true mind. In the process – called "uploading" by some and, confusingly, "downloading" by others, and "whole brain emulation" by a third group – we could *become* machines while remaining ourselves, physically transferring the structure of our minds into capacious computer programs that generate thought and the quality of minds when they are run. Uploads would live in vivid virtual realities fitted to the needs of their simulated minds, while remaining in touch with the external world.

Is this a crypto-religious hope, the much-lampooned "Rapture of the Nerds"? It does echo religious hopes of reincarnation widespread in Asian cultures, where a non-material essence slips out of an injured or aged body to enter the waiting vessel of an unborn infant. But no, the prospect of uploading has nothing significantly in common with those ancient wishful, consoling dogmas. Naturalistic materialism, the current scientific paradigm, maintains that mind is nothing other than the sublimely complex workings of the physical brain and its bodily extensions in a world of particles and force fields. If that is what we *are*, nothing prevents us from copying – mapping – our neurological complexity into some more durable, swifter material substrate.

Still, isn't this a version of the cliché from bad horror movies: a naked brain in a vat of chemical soup? Some will complain that uploading is a nightmare proposed by body-hating, frightened computer hackers, those nerdish social incompetents allegedly fleeing from sensuous reality and human warmth.

It is true that many proponents of uploading dislike the limitations and messy urgings of the body and its ancient, now often maladaptive Darwinian drives. For others, as Max More details in his chapter, what drives the interest in uploading is a desire for more life, for the greatest possible access to this beautiful and complex universe. It can't be explained away as simple hatred or fear of the flesh.

Suppose it is true that mind and passion and soul are indeed the body at work, a whirling composite of matter and force and energy, engaged with the world. As we eat, drink, and excrete, the very atoms in our cells are regularly replaced. Should we object if mind changes its location from one kind of organized and ceaselessly replaced matter to another material substrate?

It is easy to become trapped by old preconceptions. Is the mind really a machine? If being a machine suggests clockwork or even the relatively stupid computers in our smart phones (already 50 percent more powerful than the greatest supercomputers in 1976), of course not. Even these limited computers are vastly more complex than an eighteenth-century wind-up parrot, or a nineteenth-century piano driven by a paper tape. The human brain is not like a broken-down motor-mower, and nobody ever thought it was.

Uploading need not imply a world of bloated grubs lying in the dark with their brains wired to spreadsheets and simulated worlds. On the contrary: transhumanist philosopher Max More, who intends to upload when that becomes an option (and use his new freedom to explore the stars), put his own case back in the 1990s: "I'm in the gym five days a week, plus I either run or cycle. I can boast that I do 710 lbs on the leg press. No atrophied body here!" In 2013, he added with amusement, "That was 710 lbs for 8 repetitions. I'm currently doing 720 lbs for 15 reps, so I'm definitely stronger. For 8 reps, I can do something over 800 lbs" (private communication). The initial goal of uploaders would be to emulate and enhance the brain, and that requires rich connections to external reality. It calls for give and take, building from the peculiar truth that inside our porridge-like brain matter is where our selves are generated. That fact does not repudiate the body, far from it.

A quadriplegic with no access to the world other than her mouth and ears and eyes and her vivid, courageous brain *is a person*. By contrast, the superb corpse of an Olympic athlete or concert pianist with a fatal brain injury, its metabolism sustained by medical machines, is no kind of person at all, just a tragic reminder of the fallibility of life and a storehouse for luckier transplant patients.

It's worth noting that if synthetic neurons can be made half the size of the organic varieties, replacing each brain cell after copying its structure