© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

CONTENTS

INTRODUCTION: Your Visit to the Center for Mind Design 1

CHAPTER ONE: The Age of AI 9

CHAPTER TWO: The Problem of AI Consciousness 16

CHAPTER THREE: Consciousness Engineering 33

CHAPTER FOUR: How to Catch an AI Zombie: Testing for Consciousness in Machines 46

CHAPTER FIVE: Could You Merge with AI? 72

CHAPTER SIX: Getting a Mindscan 82

CHAPTER SEVEN: A Universe of Singularities 98

CHAPTER EIGHT: Is Your Mind a Software Program? 120

CONCLUSION: The Afterlife of the Brain 148

APPENDIX: Transhumanism 151

ACKNOWLEDGMENTS 153

NOTES 157

REFERENCES 165

INDEX 173

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

INTRODUCTION

It is 2045. Today, you are out shopping. Your first stop is the Center for Mind Design. As you walk in, a large menu stands before you. It lists brain enhancements with funky names. "Hive Mind" is a brain chip allowing you to experience the innermost thoughts of your loved ones. "Zen Garden" is a microchip for Zen master-level meditative states. "Human Calculator" gives you savant-level mathematical abilities. What would you select, if anything? Enhanced attention? Mozart-level musical abilities? You can order a single enhancement, or a bundle of several.

Later, you visit the android shop. It is time to buy that new android to take care of the house. The menu of AI minds is vast and varied. Some AIs have heightened perceptual skills or senses we humans lack, others have databases that span the entire Internet. You carefully select the options that suit your family. Today is a day of mind design decisions.

This book concerns the future of the mind. It's about how our understanding of ourselves, our minds, and our nature can drastically change the future, for better or for worse. Our brains evolved for specific environments and are greatly constrained by anatomy and evolution. But artificial intelligence (AI) has opened up a vast design space, offering new materials and modes of operation, as well as novel ways to explore the space at a rate much faster than biological evolution. I call this exciting new enterprise *mind design*. Mind design is a form of intelligent design, but we humans, not God, are the designers.

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

I find the prospect of mind design humbling, because frankly, we are not terribly evolved. As the alien in the Carl Sagan film *Contact* says upon first meeting a human, "You're an interesting species. An interesting mix. You're capable of such beautiful dreams, and such horrible nightmares."[1] We walk the moon, we harness the energy of the atom, yet racism, greed, and violence are still commonplace. Our social development lags behind our technological prowess.

It might seem less worrisome when, in contrast, I tell you as a philosopher that we are utterly confounded about the nature of the mind. But there is also a cost to not understanding issues in philosophy, as you'll see when you consider the two central threads of this book.

The first central thread is something quite familiar to you. It has been there throughout your life: your consciousness. Notice that as you read this, it feels like something to be you. You are having bodily sensations, you are seeing the words on the page, and so on. Consciousness is this felt quality to your mental life. Without consciousness, there would be no pain or suffering, no joy, no burning drive of curiosity, no pangs of grief. Experiences, positive or negative, simply wouldn't exist.

It is as a conscious being that you long for vacations, hikes in the woods, or spectacular meals. Because consciousness is so immediate, so familiar, it is natural that you primarily understand consciousness through your own case. After all, you don't have to read a neuroscience textbook to understand what it feels like, from the inside, to be conscious. Consciousness is essentially this kind of inner feel. It is this kernel—your conscious experience—which, I submit, is characteristic of having a mind.

Now for some bad news. The second central thread of the book is that failing to think through the philosophical

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

implications of artificial intelligence could lead to the failure of conscious beings to flourish. For if we are not careful, we may experience one or more *perverse realizations* of AI technology—situations in which AI fails to make life easier but instead leads to our own suffering or demise, or to the exploitation of other conscious beings.

Many have already discussed AI-based threats to human flourishing. The threats range from hackers shutting down the power grid to superintelligent autonomous weapons that seem right out of the movie *The Terminator*. In contrast, the issues I raise have received less attention. Yet they are no less significant. The perverse realizations I have in mind generally fall into one of the following types: (1) overlooked situations involving the creation of conscious machines and (2) scenarios that concern radical brain enhancement, such as the enhancements at the hypothetical Center for Mind Design. Let's consider each kind of scenario in turn.

CONSCIOUS MACHINES?

Suppose that we create sophisticated, general-purpose AIs: AIs that can flexibly move from one kind of intellectual task to the next and can even rival humans in their capacity to reason. Would we, in essence, be creating *conscious* machines—machines that are both selves and subjects of experience?

When it comes to how or whether we could create machine consciousness, we are in the dark. One thing is clear, however: The question of whether AIs could have experience will be key to how we value their existence. Consciousness is the philosophical cornerstone of our moral systems, being central to our judgment of whether someone or something is a self or person rather than a mere automaton. And if an AI is a conscious being,

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

forcing it to serve us would be akin to slavery. After all, would you really be comfortable giving that android shop your business if the items on the menu were conscious beings—beings with mental abilities rivaling, or even exceeding, those of an unenhanced human?

If I were an AI director at Google or Facebook, thinking of future projects, I wouldn't want the ethical muddle of inadvertently designing a conscious system. Developing a system that turns out to be conscious could lead to accusations of AI slavery and other public-relations nightmares. It could even lead to a ban on the use of AI technology in certain sectors.

I'll suggest that all this may lead AI companies to engage in *consciousness engineering*—a deliberate engineering effort to avoid building conscious AI for certain purposes, while designing conscious AIs for other situations, if appropriate. Of course, this assumes consciousness is the sort of thing that can be designed in and out of systems. Consciousness may be an inevitable by-product of building an intelligent system, or it may be altogether impossible.

In the long term, the tables may turn on humans, and the problem may not be what we could do to harm AIs, but what AI might do to harm us. Indeed, some suspect that synthetic intelligence will be the next phase in the evolution of intelligence on Earth. You and I, how we live and experience the world right now, are just an intermediate step to AI, a rung on the evolutionary ladder. For instance, Stephen Hawking, Nick Bostrom, Elon Musk, Max Tegmark, Bill Gates, and many others have raised "the control problem," the problem of how humans can control their own AI creations, if the AIs outsmart us.[2] Suppose we create an AI that has human-level intelligence. With self-improvement algorithms, and with rapid computations, it could quickly discover ways to become vastly smarter than

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

us, becoming a superintelligence—that is, an AI that outthinks us in every domain. Because it is superintelligent, we probably can't control it. It could, in principle, render us extinct. This is only one way that synthetic beings could supplant organic intelligences; alternatively, humans may merge with AI through cumulatively significant brain enhancements.

The control problem has made world news, fueled by Nick Bostrom's recent bestseller: *Superintelligence: Paths, Dangers and Strategies.*[3] What is missed, however, is that consciousness could be central to how AI values *us.* Using its own subjective experience as a springboard, superintelligent AI could recognize in us the capacity for conscious experience. After all, to the extent we value the lives of nonhuman animals, we tend to value them because we feel an affinity of consciousness—thus most of us recoil from killing a chimp, but not from eating an orange. If superintelligent machines are not conscious, either because it's impossible or because they aren't designed to be, we could be in trouble.

It is important to put these issues into an even larger, universe-wide context. In my two-year NASA project, I suggested that a similar phenomenon could be happening on other planets as well; elsewhere in the universe, other species may be outmoded by synthetic intelligences. As we search for life elsewhere, we must bear in mind that the greatest alien intelligences may be *postbiological,* being AIs that evolved from biological civilizations. And should these AIs be incapable of consciousness, as they replace biological intelligences, the universe would be emptied of these populations of conscious beings.

If AI consciousness is as significant as I claim, we'd better know if it can be built, and if we Earthlings have built it. In the coming chapters, I will explore ways to determine if synthetic

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

consciousness exists, outlining tests I've developed at the Institute for Advanced Study in Princeton.

Now let's consider the suggestion that humans should merge with AI. Suppose that you are at the Center for Mind Design. What brain enhancements would you order from the menu, if anything? You are probably already getting a sense that mind design decisions are no simple matter.

COULD YOU MERGE WITH AI?

I wouldn't be surprised if you find the idea of augmenting your brain with microchips wholly unnerving, as I do. As I write this introduction, programs on my smartphone are probably tracking my location, listening to my voice, recording the content of my web searches, and selling this information to advertisers. I think I've turned these features off, but the companies building these apps make the process so opaque that I can't be sure. If AI companies cannot even respect our privacy now, think of the potential for abuse if your innermost thoughts are encoded on microchips, perhaps even being accessible somewhere on the Internet.

But let's suppose that AI regulations improve, and our brains could be protected from hackers and corporate greed. Perhaps you will then begin to feel the pull of enhancement, as others around you appear to benefit from the technology. After all, if merging with AI leads to superintelligence and radical longevity, isn't it better than the alternative—the inevitable degeneration of the brain and body?

The idea that humans should merge with AI is very much in the air these days, being offered both as a means for humans to avoid being outmoded by AI in the workforce, and as a path

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

to superintelligence and immortality. For instance, Elon Musk recently commented that humans can escape being outmoded by AI by "having some sort of merger of biological intelligence and machine intelligence."[4] To this end, he's founded a new company, Neuralink. One of its first aims is to develop "neural lace," an injectable mesh that connects the brain directly to computers. Neural lace and other AI-based enhancements are supposed to allow data from your brain to travel wirelessly to one's digital devices or to the cloud, where massive computing power is available.

Musk's motivations may be less than purely altruistic, though. He is pushing a product line of AI enhancements, products that presumably solve a problem that the field of AI itself created. Perhaps these enhancements will turn out to be beneficial, but to see if this is the case, we will need to move beyond all the hype. Policymakers, the public, and even AI researchers themselves need a better idea of what is at stake.

For instance, if AI cannot be conscious, then if you substituted a microchip for the parts of the brain responsible for consciousness, you would end your life as a conscious being. You'd become what philosophers call a "zombie"—a nonconscious simulacrum of your earlier self. Further, even if microchips could replace parts of the brain responsible for consciousness without zombifying you, radical enhancement is still a major risk. After too many changes, the person who remains may not even be you. Each human who enhances may, unbeknownst to them, end their life in the process.

In my experience, many proponents of radical enhancement fail to appreciate that the enhanced being may not be you. They tend to sympathize with a conception of the mind that says the mind is a software program. According to them, you can enhance your brain hardware in radical ways and still run the

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

same program, so your mind still exists. Just as you can upload and download a computer file, your mind, as a program, could be uploaded to the cloud. This is a technophile's route to immortality—the mind's new "afterlife," if you will, that outlives the body. As alluring as a technological form of immortality may be, though, we'll see that this view of the mind is deeply flawed.

So, if decades from now, you stroll into a mind design center or visit an android store, remember, the AI technology you purchase could fail to do its job for deep philosophical reasons. *Buyer beware*. But before we delve further into this, you may suspect that these issues will forever remain hypothetical, for I am wrongly assuming that sophisticated AI will be developed. Why suspect any of this will happen?

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

CHAPTER ONE

THE AGE OF AI

You may not think about AI on a daily basis, but it is all around you. It's here when you do a Google search. It's here beating the world *Jeopardy!* and Go champions. And it's getting better by the minute. But we don't have general purpose AI yet—AI that is capable of holding an intelligent conversation on its own, integrating ideas on various topics, and even, perhaps, outthinking humans. This sort of AI is depicted in films like *Her* and *Ex Machina*, and it may strike you as the stuff of science fiction.

I suspect it's not that far away, though. The development of AI is driven by market forces and the defense industry—billions of dollars are now pouring into constructing smart household assistants, robot supersoldiers, and supercomputers that mimic the workings of the human brain. Indeed, the Japanese government has launched an initiative to have androids take care of the nation's elderly, in anticipation of a labor shortage.

Given the current rapid-fire pace of its development, AI may advance to artificial general intelligence (AGI) within the next several decades. AGI is intelligence that, like human intelligence, can combine insights from different topic areas and display flexibility and common sense. Indeed, AI is already projected to outmode many human professions within the next decades. According to a recent survey, for instance, the most-cited AI researchers expect AI to "carry out most human professions at least as well as a typical human" within a 50 percent probability by 2050, and within a 90 percent probability by 2070.[1]

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

I've mentioned that many observers have warned of the rise of superintelligent AI: synthetic intelligences that outthink the smartest humans in every domain, including common sense reasoning and social skills. Superintelligence could destroy us, they urge. In contrast, Ray Kurzweil, a futurist who is now a director of engineering at Google, depicts a technological utopia bringing about the end of aging, disease, poverty, and resource scarcity. Kurzweil has even discussed the potential advantages of forming friendships with personalized AI systems, like the Samantha program in the film *Her*.

THE SINGULARITY

Kurzweil and other transhumanists contend that we are fast approaching a "technological singularity," a point at which AI far surpasses human intelligence and is capable of solving problems we weren't able to solve before, with unpredictable consequences for civilization and human nature.

The idea of a singularity comes from mathematics and physics, and especially from the concept of a black hole. Black holes are "singular" objects in space and time—places where normal physical laws break down. By analogy, the technological singularity is projected to cause runaway technological growth and massive changes to civilization. The rules by which humanity has operated for thousands of years will abruptly cease to hold. All bets are off.

It may be that the technological innovations are not so rapid-fire that they lead to a full-fledged singularity in which the world changes almost overnight. But this shouldn't distract us from the larger point: we must come to grips with the likelihood that as we move further into the twenty-first century, humans may not be the most intelligent beings on

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

INDEX

Page numbers in *italics* indicate illustrations.

Aaronson, Scott, 64
ACT test, 50–57, 60, 65, 67
Active SETI, 105–9
afterlife of the brain, 8, 145
AGI (artificial general intelligence), 9, 43
AI (artificial intelligence), 1–15, 148–50
alien/extraterrestrial, 5, 98–119
  (*See also* alien/extraterrestrial AI)
    consciousness issues, 2–6 (*See also* consciousness)
    development of, 9–10
    implications, importance of thinking through, 2–3, 10
    Jetsons fallacy and, 12–13
    merging humans with AI, 6–8, 72–81 (*See also* merging humans with AI)
    mind design, concept of, 1
    postbiological, 99
    singularity, approach of, 11–12
    software, mind viewed as, 7–8, 120–47 (*See also* software, mind viewed as)
    transhumanism, 13–15 (*See also* transhumanism)
    uncertainties and unknowns regarding, 15
AI consciousness, problem of, 3–6, 16–32, 148–49
    alien intelligences, postbiological, 5
    alien/extraterrestrial AI, 110–11
    biological naturalism, arguments against, 18–22, 34, 158n4
    capability of machines for consciousness, 17–18
    Chinese Room thought experiment and, 19–22, *20*, 34, 148
    control problem, 4–5
    ethical treatment of conscious/potentially conscious AIs, 39, 67–69, 149
    isomorph thought experiment and, 26–31, 57, 158nn13–14, 159nn10–11
    "problem of other minds" and, 158n3
    slavery and, 4, 39
    techno-optimism, arguments for, 18, 23–26, 31, 34
    value placed on humans by, 5
    "Wait and See Approach" to, 33–34, 45
AI slavery, 4, 39
Alcor, *121*, 145
alien/extraterrestrial AI, 5, 98–119
    BISAs (biologically inspired superintelligent aliens), 113–19
    consciousness, 110–11
    control problem and, 104–5
    postbiological cosmos approach, 99–104
    SETI (Search for Extraterrestrial Intelligence), 101, 105–9, *106*

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

alien/extraterrestrial AI (cont.)
software theory, 119
superintelligent AI minds, encountering, 109–19
Alzheimer's disease, 44, 58
Amazon, 131
Arrival (film), 107
artificial general intelligence (AGI), 9, 43
artificial intelligence. *See* AI
asbestos, 66
Asimov, Isaac, "Robot Dreams," 57
astronauts and conscious AI, 41–43, *42*, 103

Battlestar Galactica (TV show), 99
Bello, Paul, 159n1
Berger, Theodore, 44
Bess, Michael, 12
Big Think, 126
biological naturalism, 18–22, 34, 158n4
biologically inspired superintelligent aliens (BISAs), 113–19
Black Box Problem, 46
black holes, 10
Blade Runner (film), 17, 57
Block, Ned, 159n1, 162n11
"The Mind as the Software of the Brain," 134
body. *See specific entries at* mind-body
Bostrom, Nick, 4, 77, 80, 151, 160n1
Superintelligence: Paths, Dangers, and Strategies, 5, 105, 111–13
"boxing in" AIs, 53–54, 57
Brain Preservation Foundation, 126
brain-based materialism, 76–77, 80, 160n9
Bringsjord, Selmer, 159n1
Buddha, 76, 102, 137

CAC (computationalism about consciousness), 24–25
Cartesian dualism, 140–41
Center for Mind Design (thought experiment), 1, 3, 6, 12, 70, 88, 89, 94, 148, 150
Center of Theological Inquiry, Princeton, 110
cerebellar agenesis, 22, 63
Chalmers, David, 17, 18, 139
The Conscious Mind, 28
change, self's survival during, 143–44
Chinese Room (thought experiment), 19–22, *20*, 34, 148
chip test, 44, 57–61, 65, 67
cholera, 66
Clarke, Arthur C., 119
clinicaltrials.gov, 157n4
cognitive or functional consciousness, 49–50, 51, 67
cognitive science, 23, 158n9, 161n8
collective superintelligence, 111
combinatorial language-like mental representations, 116–17
Computational Cartesianism, 140
computational structure of alien brains, 115–16
computational theory of mind, 23–25, 78–81, 126, 158n4, 161n11
connectionism, 161n11
connectomics, 115–16
consciousness, 2–6. *See also* AI consciousness, problem of; consciousness engineering; testing for consciousness in machines
concept of, 2
human moral systems and, 5, 39, 40–41, 110
mind-body problem and, 127–28

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

in new versus experienced drivers, 35–36
phenomenal versus cognitive or functional, 49–50, 51
reasons behind, 17
consciousness engineering, 4, 33–45
astronauts and conscious AI, 41–43, *42*, 103
to avoid conscious AI, 4, 39–40
consumer desire for, 41
economics and practicalities of AI production and, 37–39
human-machine mergers, 44–45
outmoding of consciousness in superintelligent AIs, 35–36
reasons for creating conscious AIs, 40–43
as safety issue, 37, 40–41
Contact (film), 2
control problem
AI consciousness and, 4–5
alien/extraterrestrial AI and, 104–5
cryonics, 120–22, *121*, 145–47
cyborgs, 12, 72
Cylon Raider, 99

DARPA (Defense Advanced Research Projects Agency), 157n4
Data from *Star Trek*, 135–37
Davies, Paul, 43
decomposition, functional, 118, 159n9
Defense Advanced Research Projects Agency (DARPA), 157n4
Descartes, René, 57, 129, 140
Dick, Steven, 101
dynamic systems theory, 161n11
Dyson, Freeman, 41
Dyson spheres, 107, *108*

essential properties, 75, 85, 142
Eternal Sunshine of the Spotless Mind (film), 90
ethics. *See also* morality
consciousness and human moral systems, 5, 39, 40–41, 110
creation of AIs with uncertain conscious status, 68–69
Precautionary Principle, 66–69
slavery and AI consciousness, 4, 39
treatment of conscious/potentially conscious AIs, 39, 67–69, 149
Ex Machina (film), 9, 149
excluded middle, principle of, 68–69
extraterrestrial intelligence. *See* alien/extraterrestrial AI
Extropians, 14

Facebook, 4, 120, 131
Francis (pope), 110
Freaky Friday (film), 51
functional decomposition, 118, 158n9
functional immortality, 8, 135, 137
functional or cognitive consciousness, 49–50, 51, 67
fusion-optimism, 73, 80–82, 87–89, 92, 95, 119, 122, 125, 133–34, 138, 163n11
Future of Humanity Institute, Oxford University, 13

Garza, Mara, 68
Gates, Bill, 4, 105

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

global workspaces, 117–18
Go (game), 9
Google, 4, 9, 10, 12, 39

HAL 9000, 53
Harmon, Amy, 122, 145
Hawking, Stephen, 4, 41, 104, 120, 122, 124
Hayworth, Ken, 126
Her (film), 9, 10, 14, 19, 37, 41, 46
Heraclitus, 143
hippocampus, artificial, 12, 44, 69, 145
Hopkins, Anthony, 33
"hot zone" of sensory processing, 38
Hughes, James, 161n10
Hume, David, 76
Huxley, Julian, 13

I, Robot (film), 149
IBM True North chip, 64
idealism, 130, 163n17
IIT (integrated information theory), 61–65
illusion, self viewed as, 76, 77, 137, 161n10
immortality, functional, 8, 135, 137
Independence Day (film), 107
instantiation of software program, mind viewed as, 134–44
Institute for Advanced Study, Princeton, 6, 41–43
instrumental convergence thesis, 113
integrated information theory (IIT), 61–65
isomorph thought experiment, 26–31, 57, 158n10, 159nn11, 13, 14

Jackson, Frank, 159n5
Jeopardy (TV show), 9
The Jetsons (TV show) and Jetsons fallacy, 12, 69

Kaku, Michio, 77
Kernel, 12, 58
Knowledge Argument, 159n5
Koch, Christof, 63–64
Kurzweil, Ray, 10, 14, 16, 33, 73, 75–76, 77, 80, 85, 143

language-like mental representations, 116–17
The Late Lessons from Early Warnings Report, 66
Locke, John, *Essay Concerning Human Understanding,* 161n8
"locked in" patients (conscious but unable to communicate), 63
Lowe, E. J., 141, 163n25

Mandik, Pete, 102–3, 119
The Matrix (film), 125
Mazie, Steven, 126
memory
  partial erasure of, 89–90, 91
  patients unable to lay down, artificial hippocampus for, 44
  reduplication problem, 84–85, 88
merging humans with AI, 6–8, 72–81. *See also* afterlife of the brain; paternism; software, mind viewed as; transhumanism
  AI consciousness problem, as way of dealing with, 6–7
  computational theory of mind and, 78–81
  consciousness engineering, as means of, 44–45

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

current efforts at, 13
fusion-optimism regarding, 73, 80–82, 87–89, 92, 95, 119, 122, 125, 133–34, 138, 163n11
personhood and personal identity, defining, 74–77, 96–97
projected trajectories of, 73–74
testing for consciousness and, 69–71
thought experiment, 72–73
zombification problem, 7
metaphysical humility, value of stance of, 94–97
Millner, Uri, 41
mind design, concept of, 1
Mind-Body Problem, 127–30, 139, 141
mind-body separation, ability to imagine, 51, 55, 57
modified patternism, 88, 89, 90, 92, 93
Moore's Law, 74
morality, 3, 31, 39, 53, 68, 105, 151. *See also* ethics
MP3 recordings, 37–38
Musk, Elon, 4, 7, 73, 77

N=1 or *small-N* approach, 63, 100
nanobots, 89, 92, 95, 144
NASA, 5, 101
neural correlate for consciousness, 60–61, *61*
neural lace, 7, 12, 69
Neuralink, 12, 58
neuromorphic systems, 40, 64, 158n4
Nietzsche, Friedrich, 76, 137
Thus Spoke Zarathustra, 82
no-self view, 76, 77, 137, 161n10
Olson, Eric, 161n9
orthogonality thesis, 112

panpsychism, 63, 128, 130, 139, 158n7, 163n17
Parfit, Derek, 96, 102, 137
patternism, 77, 82–97
change, self's survival during, 143–44
enhancements through gradual replacement of neural tissue, 88–92, 95
metaphysical humility, value of stance of, 94–97
modified patternism, 88, 89, 90, 92, 93
personhood and personal identity, defining, 96–97
reduplication problem, 84–85, 88, 123–24
Sawyer's *Mindscan* and uploading minds, 82–84, 86–87, 95
software approach to the mind and, 78–81, 125, 127, 134, 143–44
spatiotemporal continuity and, 85–89, *86–87,* 93, 127, 163n10
substrates, transferring, 92–94
theory of, 78–81
PC (phenomenal consciousness), 49, 50, 51, 56
Pearce, David, 151
personhood and personal identity, defining, 74–77, 96–97
phenomenal consciousness (PC), 49, 50, 51, 56
Φ values, 62–65, 160n10

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

physicalism, 76, 129, 139, 158n4, 161n9
postbiological cosmos approach, 99–104
postbiological intelligence, 99. *See also* AI
posthumans, 72, 73, 86, 143, 152
post-traumatic stress disorder (PTSD), 44, 58
Precautionary Principle, 66–69
"problem of other minds," 157n3
program instantiation, mind viewed as, 134–44
Program View of the mind. *See* software, mind viewed as
property dualism, 129–30, 139
psychological continuity theory, 76, 77, 89, 160n8
PTSD (post-traumatic stress disorder), 44, 58
Putnam, Hilary, 140

quality of consciousness, judging, 48–49
quality superintelligence, 112
Quantum Information Center, University of Texas at Austin, 64

R2D2, 33, 126
reduplication problem, 84–85, 123–24
religion, AIs, and extraterrestrials, 110
Rucker, Rudy, *Software,* 125

Sacks, Oliver, 44
safety issue, consciousness engineering as, 37, 40–41
Sagan, Carl, 2, 98
SAM (software approach to the mind), 124–27. *See also* software, mind viewed as
Sawyer, Robert, *Mindscan,* 82–84, 86–87, 95
Scharf, Caleb, 41–42
Schisler, Josh, 120–21, 145
Schneider, Susan
  The Language of Thought, 126
  at Starshot Initiative, 41–43
Schwitzgebel, Eric, 68
Search for Extraterrestrial Intelligence (SETI), 101, 105–9, *106*
Searle, John, 19–22, *20,* 158n4
self viewed as illusion, 76, 77, 137, 161n10
sensory processing "hot zone," 38
separation of mind from body, ability to imagine, 51, 55, 57
SETI (Search for Extraterrestrial Intelligence), 101, 105–9, *106*
SIM and SIM* (software instantiation view of/approach to the mind), 134–42. *See also* software, mind viewed as
singularity, approach of, 10–11
Siri, 37
slavery and AI consciousness, 4, 39
small-*N* or *N*=*1* approach, 63, 100
social development, human, versus technological progress, 2
software, mind viewed as, 7–8, 120–47
  abstract nature of programs and, 131–33, *132*
  alien/extraterrestrial AI and, 119
  change, self's survival of, 143–44
  computational theory of mind, 23–25, 78–81, 158n4
  functional immortality and, 135, 137

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

merging humans with AI and, 78–81
Mind-Body Problem and, 127–30, 139, 141
nonviability of, 126, 130–34
patternism and, 78–81, 125, 127, 134, 143–44
SAM (software approach to the mind), 124–27
SIM and SIM* (software instantiation view of/approach to the mind), 134–44
Suozzi, Kim, and, 120–22, *121*, 133, 138, 144–46, 145
uploading and, 122–26, 133, 136–37, 146–47
soul theory, 76, 77, 88
spatiotemporal continuity, 85–89, *86–87*, 93, 127, 163n10
speed superintelligence, 111
Star Trek: The Next Generation (TV show), 135–37
Star Wars (film), 12–13, 33
Starshot Initiative, 41–43, *42*
substance dualism, 128–29, 140–41
substrate independence and transferring substrates, 24, 92–94, 150
Summit supercomputer, 11–12
Suozzi, Kim, 120–22, *121*, 133, 138, 144–46, 145
superintelligence, 35–36, 109–19
symbolic/language of thought approach, 161n11
Synapse (U.S. Defense Department), 14
Systems Reply to Chinese Room conundrum, 21–22

technological progress versus human social development, 2
techno-optimism
about AI consciousness, 18, 23–26, 31, 34
merging with AI and, 73
Tegmark, Max, 4
Terminator films, 3, 104
testing for consciousness in machines, 5–6, 46–71
ACT test, 50–57, 60, 65, 67
chip test, 44, 57–61, 65, 67
difficulties and complications in, 46–51
IIT (integrated information theory) and, 61–65
mind-machine mergers and, 69–71
responding when machines test positive for consciousness, 65–69
separation of mind from body, ability to imagine, 51, 55, 57
Turing test, 56
theory of mind, 58n3
Tononi, Giulio, 61–64
Transcendence (film), 124–25
transhumanism, 13–15, 151–52. *See also* merging humans with AI
on AI consciousness, 16
defined, 73
enhancements rejected by, 160n1
patternism and, 77–81
World Transhumanist Association, 151
Transhumanist Declaration, 14, 151–52
"The Transhumanist Frequently Asked Questions," 80, 95–96, 152
TrueNorth chip (IBM), 64
Turing, Alan, and Turing test, 56, 140

For general queries, contact webmaster@press.princeton.edu

© Copyright, Princeton University Press. No part of this book may be distributed, posted, or reproduced in any form by digital or mechanical means without prior written permission of the publisher.

Turner, Edwin, 41–43, 54
2001: A Space Odyssey (film), 53

UNESCO/COMEST report on Precautionary Principle, 66
uploading
  patternism and, 80–81, 82–84, 95
  software, mind viewed as, 122–26, 133, 136–37, 146–47

vegetative state, human patients in, 61–62

Westworld (TV show), 17, 33, 45
Witkowski, Olaf, 41–42
World Transhumanist Association, 151

X-files (TV show), 116

zombies, 7, 41, 49–50, 51, 56, 88, 102, 131

For general queries, contact webmaster@press.princeton.edu