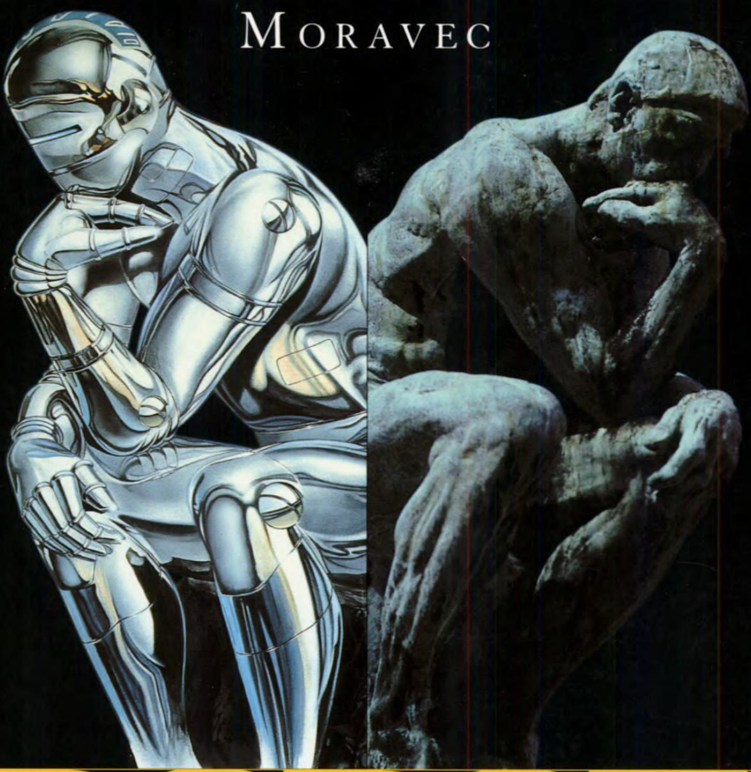
HANS MORAVEC



ROBOT

MERE MACHINE *to* TRANSCENDENT MIND

Robot

Mere Machine to Transcendent Mind

Hans Moravec

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford New York
Athens Auckland Bangkok Bogotá Buenos Aires Calcutta
Cape Town Chennai Dar es Salaam Delhi Florence Hong Kong Istanbul
Karachi Kuala Lumpur Madrid Melbourne Mexico City Mumbai
Nairobi Paris São Paulo Singapore Taipei Tokyo Toronto Warsaw

and associated companies in

Berlin Ibadan

First published by Oxford University Press, Inc., 1999

First issued as an Oxford University Press paperback, 2000
198 Madison Avenue, New York, New York 10016

Library of Congress Cataloging-in-Publication Data
Moravec, Hans P.
Robot: mere machine to transcendent mind /
by Hans Moravec.
p. cm.
ISBN 0-19-511630-5 (Cloth)
ISBN 0-19-513630-6 (Pbk.)
1. Robotics I. Title
TJ211.M655 1999 303.48'34'0112--dc21 97--47328

3 5 7 9 10 8 6 4 2
Printed in the United States of America
on acid-free paper

1

Escape Velocity

Progressive change sculpted our universe and our societies, but only very recently has human culture seen beyond the short cycles of day and night, summer and winter, birth and death, to recognize it.[1] No sooner was universal change noted in the traces of history than its accelerating pace became discernible in a single lifetime. By almost any measure—energy, information, speed, distance, temperature, variety—the developed world is growing more capable and complex faster than ever before—a statement that has been true for at least half a millennium, and mostly true since the agricultural revolution and the invention of writing over five thousand years ago.

Many of the products of this accelerating process—written language, city-states, and automation, for instance—sped it further. Today the pace strains the limits of human adaptability: the lessons of a technical education are often obsolete before the education is complete. Nevertheless, the acceleration continues, as machines take over where humans falter. In the 1970s photographic patterns for manufacturing integrated circuits with dozens or hundreds of components were designed and drafted by hand, on plastic sheets. Today's computer chips contain tens of millions of components, placed by design programs running on older computers. Not only did one generation of machines make possible the next, they enabled it to appear in less than a year, compared to an average of three years for purely human designs.

Self-accelerated computer evolution affects other technical fields, and almost every design engineer's work is being amplified and accelerated by computer workstations and communications. The many parts of the new Boeing 777 aircraft, to pick a major example, were designed in parallel, by distantly separated engineering teams, with powerful three-dimensional modeling programs. Subassembly designs were checked for compatibility by programs that put together the aircraft in simulation, detecting major and minor mechanical, electrical, control, and aerodynamic problems while they were easy to correct, long before a physical prototype was built. The result was an aircraft of unprecedented complexity brought into existence in half the time of previous models. In the same way, chemists and biologists are replacing years of wet lab work with weeks of molecular simulation. Architects quadrupled their business capacity by replacing drafting boards and manual bookkeeping with computer workstations in the early 1990s.

Vanishing Verities

"Things tossed up come down" is an early theory of gravity, demonstrably true in everyday life, unquestioned for millennia, until Newton developed a new theory of gravity that gave stable orbits to sufficiently fast satellites, and let slightly faster projectiles escape to infinity. "Wood rubbed warm cools down" may have been a truism for our distant ancestors, until one of them rubbed hard enough to achieve ignition temperature, whereupon the wood flamed hotter than ever on its own. "Machines break down" is a demonstrable truth of industrial society, but as machines increasingly design, diagnose, and repair themselves, it too will be suddenly invalidated. Once given "escape velocity," machines more capable than any we know will, without further help from us, grow more capable still, learning from the world, as we did in our biological and cultural evolution. The wood is already smoldering.

Like passengers in a rising elevator, those riding a developmental curve may be unaware of the altitude already reached—until a passing window shows a glimpse of the ground. In 1930 an Australian gold-prospecting party flew into a supposedly uninhabited area deep in the New Guinea highlands and encountered a human culture separated fifty thousand years from their own. The naked inhabitants, some with stone spears, were driven into paroxysms of confusion and religious fear and awe by the giant roaring silver birds that alighted near their mud-thatch villages to release droopy-skinned white men without genitals who, among too many wonders, captured their souls in small black boxes labeled Kodak.[2]

In 1991 Davi Kopenawa took the giant step of being the first to leave the jungle to speak for his people, the Amazonian Yanomami. The Yanomami, with a population of about twenty thousand the largest remaining stone-age tribe, were isolated from the rest of the world for ten thousand years until this century, when missionaries, anthropologists, and, more recently, highway workers and gold miners, began to invade their homeland. Accompanied by a translator, wearing his only possessions, sneakers, jeans, and a sweater given to him for the trip, he visited New York, Washington, and Pittsburgh, to beg to be left alone: foreign diseases, especially malaria, had killed one-fifth of the Brazilian Yanomami in five years.

What he saw in the cities horrified him: crazy ant-people crawling in sky-high huts thinking about cars, money, and possessions instead of relatives and nature. In a zoo he identified with the listless animals among plastic plants, steel vines, and bad air. "If I had to live in your cities for a month, I'd die. There's no forest here."

Kopenawa has a point. The world we inhabit is radically different, culturally and physically, from the one to which we adapted biologically. We were shaped during the last two million years by an ongoing Ice Age—a time of continuous climatic change, as every few tens of thousands of years glaciers advanced and retreated over most of the earth (the current warm spell is but an interglacial period). Such variability favors high adaptability, by making life untenable for the rigidly optimized. In our species the adaptability took the form of a hypertrophied brain and an extended childhood, supporting an extreme cultural plasticity, along with an ever more expressive language to rapidly pass on adopted behaviors: as we grow to puberty we can learn equally well to be fur-clad arctic hunters, robed desert nomads, or naked equatorial gatherers. For almost all of human history, as still in Kopenawa's world, cultural inheritance played a straightforward supporting role: providing the *how* for the basic needs of life. But somewhere, about five thousand years ago in our cultural history, the relationship between biology and culture began to alter radically.

The Cultural Revolution

Culture lets us rapidly accommodate to environmental changes because it is a medium for a new kind of evolution. Collections of rules for behavior (memes, to use a term invented by Richard Dawkins[3]) pass from generation to generation, mutating and competing with alternatives, just as biological genes do—only much more quickly. A biological trait requires generations of selective replication to become established in a population, but a cultural practice can be altered, and spread through an entire tribe, many times in a single human lifetime. After hundreds of thousands of years of slow cultural meander, our ancestors stumbled into a set of behaviors that catalyzed the creation of ever more behaviors and memories, and physical implements to support them: a self-accelerating cycle that is reaching escape velocity today. What exactly sparked the tinder, apart from simple accumulation of useful skills, is a fascinating question. A baby boom or forced migration in an improving climate may have led to shortages in hunting and gathering resources, forcing would-be survivors into agricultural life, and eventually the first agricultural civilizations, in the Near East and China ten millennia ago.

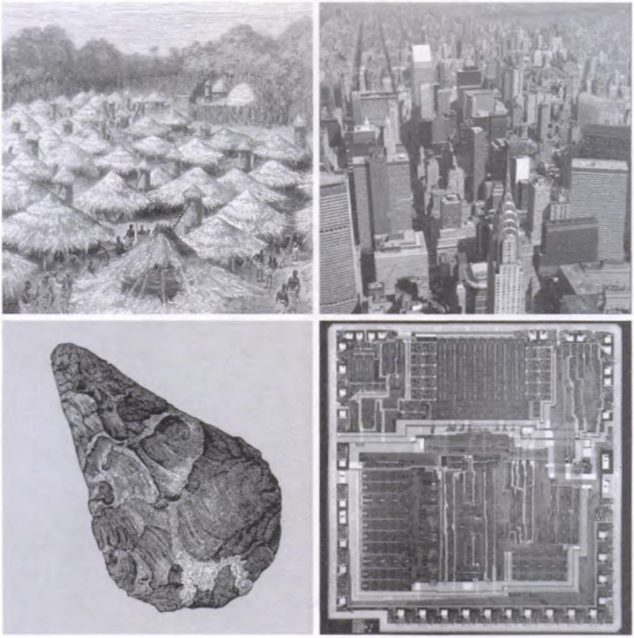
For millions of years, primates, our ancestors included, have lived in tribes. Among primates (as in canine packs, but unlike herd animals) individuals know one other personally, and maintain long-term one-on-one relationships, involving dominance, submission, friendship, enmity, debts, grudges, and intrigues—the stuff of soap opera. Complex socialization gives the tribe great abilities. In critical circumstances, individuals know who to trust with what tasks. But remembering many things about many individuals ought to take storage space in the brain. Robin Dunbar[4] has indeed found a tightly correlated linear relation between brain size and troop size in monkeys and apes—macaque monkeys, for instance, form bands of about fifteen, while larger-brained chimpanzees and gorillas live in tribes of thirty to forty members. This soap-opera connection likely drove the evolution towards large brains in primates, since tribes compete for food and shelter-providing territory, and a coordinated larger group is likely to beat out a smaller one, giving large tribes, and thus large brains, an advantage. Dunbar extrapolates the primate group / brain ratio curve to human brain size, and finds our natural tribes should have about two hundred individuals. In fact, this is just the maximum size of self-contained nonhierarchical human groups: Yanomami Indian villages and gypsy bands, for instance, and perhaps hippie communes. Modern society's overlapping webs of individual acquaintances muddle but don't eliminate our tribal limitations, evidenced in ubiquitous anecdotes. My wife, involved in many church organizations, notes that growing churches have major crises of identity when their membership reaches about two hundred. The computer science department at Carnegie Mellon University was known for its cooperative, "family" atmosphere in the 1970s, when it numbered about a hundred. It grew rapidly in the 1980s, and in the 1990s the over six hundred members of the School of Computer Science are divided into several departments and projects that are strangers to one another.

The agricultural civilizations were able to grow far beyond village size because of a series of social inventions, among them institutional roles like King, Soldier, Priest, Merchant, Tax Collector, and Peasant, clearly marked by costumery, ceremony, and standard rituals, substituting for the impossible task of remembering thousands of individual relationships. New solutions bring new problems. Cheaters in villages are easily recognized and punished but find many opportunities and hiding places in the anonymity of large society. Enforcement institutions—Moral authorities, Lawmakers, Police, and Criminal labels—partially countered the breakdown of cooperation. The problem of keeping track of who owes what to whom, a matter of memory in a village but a criminal opportunity in a city, encouraged the invention of recordkeeping: tokens, tally marks, a number system, and eventually writing. The new social functions involved complicated procedures unlike those of tribal life, many thus slow and difficult to learn. Enter extended formal training periods, eventually Teachers and Schools.

Like villages, civilizations compete with one another for resources and may gain advantage from institutions that foster

Community life and chipped silicon

For over a million years we evolved biologically into circumstances resembling the images on the left. The bizarre variants on the right evolved culturally, much too fast for our biology to have kept up. It is a credit to the flexibility of our ancestral design that many of us manage to squeeze into the strange new molds. It is no surprise if we find the fit uncomfortable—or that some of us never manage it at all.



innovations—and incidentally put cultural evolution into higher gear. Agriculture benefits from precise knowledge of the season, and thus of celestial cycles, and military and civil projects go better with professional thinkers and builders on the job, so the positions of Astronomer/Astrologer, Philosopher/Magician, and Engineer/Artisan become part of the picture. The innovations of professional thinkers, transmitted by increasingly effective written language over huge distances and times, accelerate innovation itself. The result is a process far, far faster than biological evolution that produces ever more elaborate places for humans to live, ever swifter ways for them to move and to communicate, ever larger storehouses of previous thought, ever more territory occupied, ever more energy controlled. It also produces a world increasingly unlike the villages, fixed and nomadic, in which human behavior evolved, and so makes ever greater demands on our adaptability.

Strange Ducks, Out of Water

Today, as our machines approach human competence across the board, our stone-age biology and our information-age lives grow ever more mismatched. Work in the developed countries has become increasingly specialized and esoteric, and it now often takes a graduate degree, representing half a working lifetime of sustained learning, to master the necessary unnatural skills. As societal roles become yet more complex, specialized, and far removed from our inborn predispositions, they require increasing years of rehearsal to master, while providing fewer visceral rewards. The essential functions of a technical society elude the understanding of an increasing fraction of the population. Even the most successful individuals often find their work boring, difficult, unnatural, and unsatisfying, more like a sustained circus performance than a real life. Caffeine substitutes for natural adrenaline. Those original activities that do remain—eating and child raising, for instance—are often squeezed by the strange new tasks. The mismatch between instinct and necessity induces alienation in the midst of unprecedented physical plenty.

By the standards of our inherited tribal psychology, we *are* sick and crazy. Physically, however, we are healthier and live longer than ever, and we have vastly more options in every sphere of activity. Few city-dwellers would be prepared to adopt the circumscribed life in a stone-age forest village, despite uneasiness with their own. On the contrary, much of the third world is rushing to overcome its physical problems by adopting the patterns of the developed nations (Davi Kopenawa himself is now a regular speaker at ecological meetings around the world, and his stories and commentary are extensively represented on the World Wide Web!). The urbanized, meanwhile, have devised substitutes for some tribal experiences, for instance, churches and other social organizations that bring together village-sized groups with a common sense of purpose, a shared experience, a defining mythology, and uniform behavioral expectations. Others find release in competitive sports (very like tribal wars), outdoor vacations, or even backyard barbecues. Some business trips resemble mammoth-hunting forays but lack the scenery, quiet stalks, and satisfying physical marksmanship—and a golfing weekend fills the void. But, as the pace, diversity, and global geographic interconnectedness of life continues to increase, even such occasional imitations of our ancestors' lifestyles are crowded out and may be becoming less satisfying. The world is rushing away from our ancestral roots ever faster, stretching the limits of both our biological and institutional adaptability.

Some individuals and communities have tried to isolate themselves from the problem. The Pennsylvania Amish live in a perpetual state of early-nineteenth-century rural industrialization. Some cloistered religious orders operate like isolated tribes. Countercultural rural communes of the 1960s and 1970s deliberately resembled villages. Yet, industrialized society's increasing population, accessibility, and competitive vigor in all fields seems to erode such communities, who cannot reasonably, or legally, deny members in need the benefits of modern medicine, inexpensive food, clothing, building materials, useful machinery, and especially empowering, but distracting, education.

There are unhappy voices today calling for a worldwide rollback to an earlier state of affairs. They are outvoted by the demands of billions for food, housing, and civilized comforts. Yet, paradoxically, as our cultural artifacts achieve self-sustaining maturity, they will provide the means to restore humanity and nature to an imitation of the wild past.

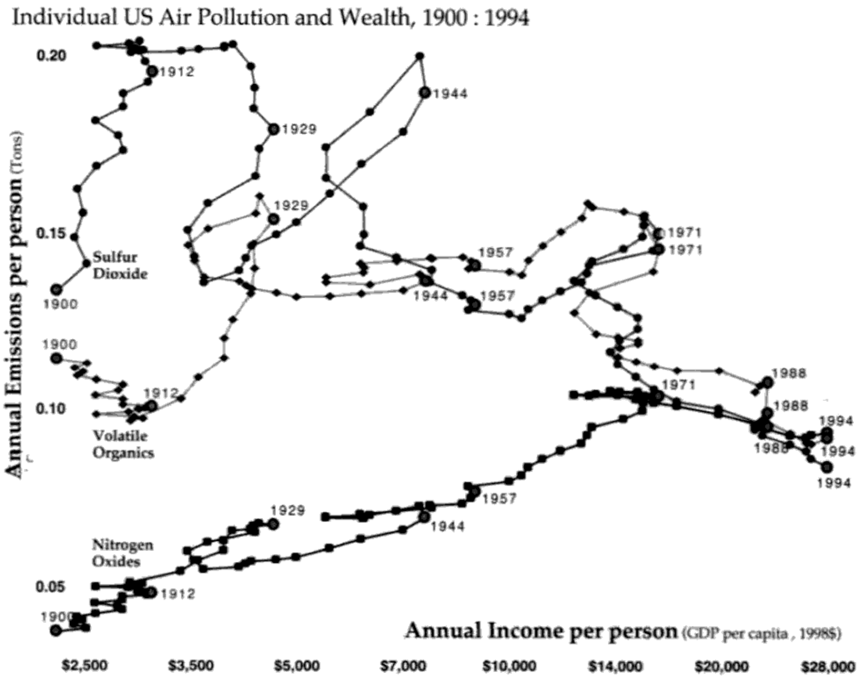
Back to the Future

Productivity rose during the Industrial Revolution, as steadily improving machines outperformed and displaced ever more human labor. Simple diffusion, and social innovations like labor unions and profit taxes, widely distributed the consequent wealth. The wealth expressed itself both in increased public and private consumption and increased leisure time. During the last three centuries in the industrialized countries, slave and child labor and hundred-hour factory work weeks have given way to under-forty-hour weeks and mandatory retirement.

Short-term fluctuations in the trend notwithstanding, as machines assume more—eventually all—of essential production, humans everywhere will be left with the options of the idle rich. Work time is yoked to the strange needs of the enterprise, but idle time can be structured to satisfy hunter-gatherer instincts. The human population will regain the opportunity to organize its life in more natural patterns. A greener planet is a likely result of this ongoing process. As societies industrialize and become wealthy, increased consumption manifests itself in deforestation, pollution, and the like—to a point.

Further wealth reduces the manifestations of industry, by making the luxury of a greener environment affordable. Advancing technology widens the options from which individuals sculpt first their personal lives, but then also their communal world. The developed countries of America, Asia, and Europe began their green return in recent decades, as per-capita annual income grew beyond about $15,000. Many developing countries are just reaching this turnaround point. Advanced robots will reinforce the trend indirectly, by tremendously accelerating technological evolution and, for instance, allowing extreme processes to be moved to outer space. They will contribute directly by substituting for energy- and chemical-intensive industrial separation and shaping processes. A robot population far exceeding the human one will achieve the same end much more efficiently by tirelessly sorting and rearrang-

Up, then down
Growing knowledge and wealth increase the options available to individuals. Personal needs and wants outweigh communal concerns, but eventually even communal goods become affordable. Despite increasing population and personal consumption, U.S. environmental pollution has mostly improved since 1970, when per capita income exceeded $15,000 in 1998 currency. A consensus of opinion set the direction, technology and wealth provided the means. (Source: U.S. Statistical Abstract of the United States, 1996, *CDROM, U.S. Census Bureau, 1997, esp. table 374, spreadsheet version. Supplemental data from* Historical Statistics of the United States: Colonial Times to 1970, *U.S. Census Bureau, 1976.)*



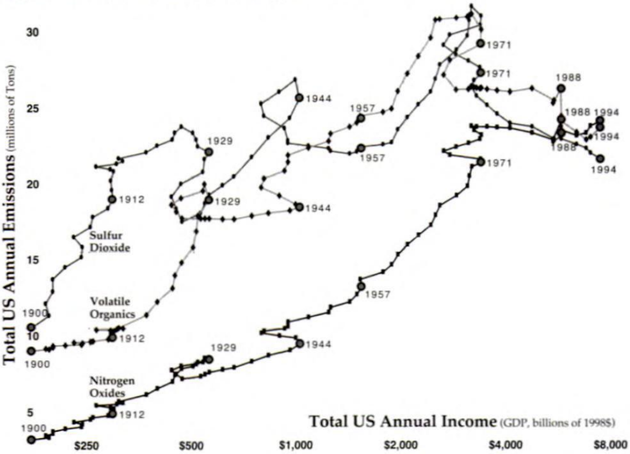
ing matter on a tiny scale with myriads of microscopic fingers.

Any choice has consequences: by comfortably retreating to its roots, biological humanity will leave the uncomfortable, uncharted but grand opportunities to others.

No less than today's organizations, fully automated companies will compete with one another not only in routine manufacture and distribution, but also in planning, development, and research. To robots built for it, outer space will offer unprecedented energy, materials, room, and perhaps freedom from taxation

Up, then down (cont)

Overall US Air Pollution and Wealth, 1900 : 1994



for these activities—a tremendous competitive advantage. Sooner rather than later, automated industry will grow away from earth. The space industries will continuously devise their own improvements, gaining rapidly in size, efficiency, diversity, and intelligence. The earthbound "consumer outlet" parts of the operation, while not shrinking in absolute size, will represent an ever-decreasing fraction of the whole. Old earth will become insignificant on the ever-grander scale of earth-spawned activity.

Robot industries will start as conversions of existing enterprises, retaining their institutional, legal, and competitive structures. But then they will explore and exploit expanding non-traditional options, some very unhuman. Our artificial progeny will grow away from and beyond us, both in physical distance and structure, and in similarity of thought and motive. In time their activities may become incompatible with old earth's continued existence. Even so, it is likely that we, the historical root of their transcendence, will be preserved in some form—though, to us, the form may seem extremely strange. Just possibly, human personalities could participate in some way in the mainstream of this future activity, either under the wings of superintelligent hosts, or by being transformed into a compatible form—surely becoming very unhuman in the process.

There is an analogy between the evolution of the first living organisms from simpler chemical processes several billion years ago and the development of technical civilization from human manipulative and learning skills. Technical civilization, and the human minds that support it, are the first feeble stirrings of a radically new form of existence, one as different from life as life is from simple chemistry. Call the new arrangement Mind. Unlike Life alone, which learns from its past, but is blind to its future, Mind can choose among alternatives to imperfectly select its own destiny—even to amplify that very ability.

Mind Fire

Chapter 2 reviews the state of the robot art, like a baby poised for sudden growth. The following chapters mix predictions with suggested actions. One of the lessons of chaos theory is that sensitive systems are impossible to predict but often easy to control. Under that model, the future *can* sometimes be predicted, if one steadily nudges events toward the prediction! Believable and physically possible predictions can themselves nudge, by inspiring work. When such proactive predictions miss, it is often because they overlook even more potent possibilities, rather than because they are unachievable.

In the thirteenth century Roger Bacon imagined high-speed worldwide travel—via seven-league boots, rather than flying conveyances. In the sixteenth century Leonardo da Vinci designed aircraft—powered by human muscle, rather than combustion engines. In the nineteenth century Jules Verne anticipated submarine warfare—against wooden sailing vessels, rather than armored battle fleets guarded by electronic senses and aircraft. Shortly thereafter, H. G. Wells anticipated a world of the distant future with humanity radically transformed—by Darwinian evolution, not directed engineering. Science fiction of the early twentieth century, inspired by the theories, inventions, and speculations of rocket pioneers like Konstantin Tsiolkovsky, Robert Goddard, and Hermann Oberth, is filled with spacecraft—guided by slide-rule-wielding human navigators, not digital computers (telephone, radio, and computers, in particular, and their dramatic applications, seem to have taken prognosticators by surprise). There are no large fleets of dirigible airships ferrying transatlantic passengers; faster and more manageable heavier-than-air craft displaced them.

Barring cataclysms, I consider the development of intelligent machines a near-term inevitability. Chapters 3 and 4 offer a scenario. Like airplanes, but unlike spaceships or radio, machine intelligences will be direct imitations of something already existing biologically. Every technical step toward intelligent robots has a rough evolutionary counterpart, and each is likely to benefit its creators, manufacturers, and users. Each advance will provide intellectual rewards, competitive advantages, and increased wealth and options of all kinds. Each can make the world a nicer place to live. At the same time, by performing better and cheaper, the robots will displace humans from essential roles. Rather quickly, they could displace us from existence. I'm not as alarmed as many by the latter possibility, since I consider these future machines our progeny, "mind children" built in our image and likeness, ourselves in more potent form. Like biological children of previous generations, they will embody humanity's best chance for a long-term future. It behooves us to give them every advantage and to bow out when we can no longer contribute.

But, as also with biological children, we can probably arrange for a comfortable retirement before we fade away. Some biological children can be convinced to care for elderly parents. Similarly, "tame" superintelligences could be created and induced to protect and support us, for a while. Such relationships require advance planning and diligent maintenance. Chapter 5 offers suggestions.

It is the "wild" intelligences, however, those beyond our constraints, to whom the future belongs. The available tools for peeking into that strange future—extrapolation, analogy, abstraction, and reason—are, of course, totally inadequate. Yet, even they suggest surreal happenings. Chapter 6's robots sweep into space in a wave of colonization, but their wake converts everything into increasingly pure thinking stuff. A "Mind Fire" will burn across the universe. Inside the Mind, considered in Chapter 7, physical law loses its primacy to purposes, goals, interpretations, and God knows what else.

2

Caution! Robot Vehicle!

A strange sight greeted motorists driving along eucalyptus-lined Arastradero Road in the foothills behind Stanford University during the 1970s. Amid rolling terrain populated by horse-riding farms, dairy herds, and an occasional biker café, the highest hilltop was crowned by a large circular structure with brightly glowing windows. A recently landed starship? An orange road sign on the steep driveway winding up the hill warned **CAUTION ROBOT VEHICLE**. Perhaps Gort, the ominous extraterrestrial robot from the 1951 movie *The Day the Earth Stood Still,* guarded the entry ramp?

The guess would not be entirely wrong. The building, on property leased from Stanford University, was to be a central research lab for General Telephone and Electric. Plans changed, and it was abandoned in mid construction in the mid-1960s. Finding no other buyers, it was acquired by the university and used transiently for storage and research groups outgrowing campus locations. John McCarthy's burgeoning "Artificial Intelligence Project"—later renamed Stanford AI Laboratory, or SAIL—resided there longest, from 1968 until 1979, to the discomfort of the Stanford administration, who found the isolated and overlarge building expensive to operate and maintain.

The road sign warned of a spindly TV-transmitting remote-controlled electric vehicle the size of a card table, known as the Stanford Cart, that sometimes ran on its four small bicycle tires at walking speed around the building's driveway and parking lot. Often it was remote-controlled by a human in the building, monitoring the robot's TV broadcast. Once in a while, the TV images were pro-

Cart under SAIL
As both robot vehicle and road hazard, the dimly aware Stanford Cart traverses the driveway of the Stanford Artificial Intelligence Laboratory in a composite image recreating the situation in the 1970s.

cessed and the robot controlled by the AI Lab's room-sized computers, and then the vehicle truly was an autonomous mobile robot. In 1971, in the Ph.D. work of a student named Rod Schmidt, the Cart, in a kind of robotic sobriety test, was programmed to slowly follow a white line. In my own studies, in 1975 it drove in straight lines by keeping its eye on a skyline of trees, and in 1979 peered and planned its way through obstacle courses, crossing a thirty-meter room in the astonishing time of five hours, getting lost about one crossing in four.

Embarrassingly subhuman performance has been typical of robotics, to the chagrin of its enthusiasts. A 1969 essay by John McCarthy, *Computer-Controlled Cars,* suggested that a medium computer of the day could drive a seeing robot car evolved from the Stanford Cart in normal traffic. More conservative visionaries thought automated traffic might soon travel on special highway lanes marked by buried signal-emitting wires. Three decades later, computers grown a hundred thousand times as powerful still can't be trusted to drive vehicles unattended, except for a few people movers that run on sensor-studded tracks.

Why was the intuition of some very clever people so far off the mark? Distances can be deceiving when one sees a territory from only one special vantage point. Distance, in this metaphor, is the absolute difficulty of accomplishing various mental, sensory, and motor feats. The special vantage point is our large brain, sculpted over hundreds of millions of years. During that time our animal ancestors were selected for their proficiency in obtaining the essentials of life and reproduction, and winning escalating arms races with competitors and parasites. In contrast, writing, arithmetic, and logical reasoning are recent cultural artifices. We find sensing and moving natural and easy, while "brain work" is tedious and difficult. An effortless glance reveals the pieces on a chessboard, but it takes long, hard thought to choose good chess moves.

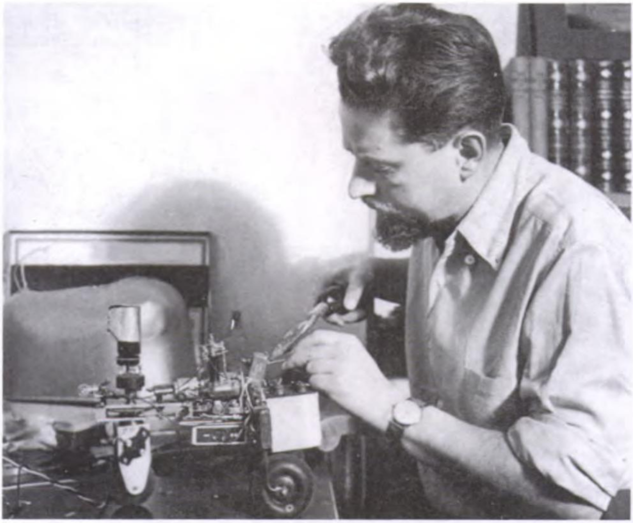
A decades-long effort to build machines that sense, act, and think has established a very different viewpoint. Robot sensing and moving, no less than thinking, is built up of large numbers of simple steps. It has become clear that reliably locating chess pieces in camera images demands thousands of times as many steps as planning good chess moves.

The serious possibility of machines behaving like animals and humans arose from the technology of World War II. On the one hand, servomechanisms, in sensors and weapons, allowed electronically driven motors to follow precise motions and respond to subtle sensory cues. On the other, digital computers, breaking codes, calculating artillery tables, and simulating atomic explosions, executed enormously long arithmetical and logical procedures with superhuman speed and accuracy.

Cybernetic Creatures

Norbert Wiener, a mathematician who had developed the theory for complex predictive bombsights and anti-aircraft gun directors, attracted a small community of psychologists, biologists, engineers, and scientists of other disciplines, mostly in America and Britain, into a new field he called *cybernetics*—the science of control and communication in the animal and the machine.[5] Under that banner, researchers organized shoeboxes of electronics into artificial nervous systems that could learn to recognize simple patterns, and turtle-like robots that avoided obstacles and sought out lights.

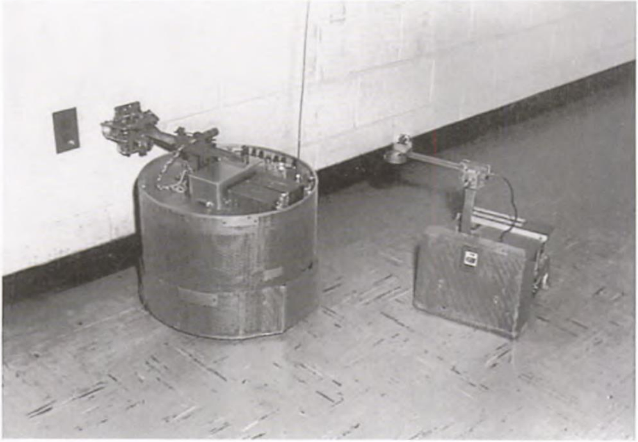
1950—Dr. W. Grey Walter operates on one of his light-seeking tortoises
When "Elsie" is fully assembled, the bump-sensing switch at the top of the mechanism supports a protective shell, seen in the background.



In England from 1948 to 1951, W. Grey Walter, a British psychologist, built a half-dozen electronic tortoises with subminiature radio-tube brains, rotating phototube eyes, and contact-switch feelers.[6] They could avoid trouble while wandering about and return to their carrying hutch when its light beckoned. In groups they exhibited unexpected social behavior (dancing) by responding to one another's control lights and touches.

In the early 1960s the availability of transistors, much smaller, cheaper and less power consuming than tubes, allowed a group of brain researchers at the Johns Hopkins University Applied Physics Laboratory (APL) to build more complex robots. The "Hopkins Beast" wandered the halls of APL, keeping itself on the midline

1964—Two early versions of the "Hopkins Beast" feeding at a wall outlet
These machines wandered hallways guided by sonar and found sockets by feeling along walls. In subsequent work, the larger unit was given another layer containing a photocell circuit that allowed it to find contrasting wall plugs optically from a distance.



by side-directed sonar. When its batteries ran low, a specialized photocell eye searched the white-painted cinder-block walls for the distinctive black cover plate of wall outlets, where the robot would plug in to feed. The Beast inspired several imitators. Some used television cameras instead of photocells, and were controlled by assemblies of transistor logic gates, like those that can now be found, in the millions, in the integrated circuits of every computer. Some added new motions such as "Shake to untangle recharging arm" to the repertoire of basic actions.

Cybernetics' early results were intriguing enough to sustain two decades of research, but the field stumbled in the late 1960s when its methods proved unsuccessful when applied to more challenging problems, such as the effort to build practical reading machines.

Artificial Intellects

Electronic computers fueled an entirely different approach to thinking machines. In 1936 Alan Turing invented the mathematical concept of a universal computer that could be programmed to imitate the action of any other information processor; he used it to show that unsolvable problems exist. During World War II, in deepest British secrecy, he put his ideas into practice in the Colossus electronic computers. Those machines broke the German U-boat codes, allowing allied shipping to evade the wolf packs and win the war. After the war, he speculated that the imitation repertoire of a universal computer included all the functions of the human brain. Although his wartime accomplishments remained a state secret until 1974, his speculations on intelligent machinery became public in a series of debates on BBC radio in 1950 and in a following article.

The themes sounded on both sides in those debates still echo, though Turing died in 1953. In the United States, John von Neumann, who had developed mathematics and computers used in the design of atom bombs, was also speculating on artificial imitations of life and thought. His investigation, too, was cut short by death in 1956. But the question was in the air: with clever programming, might not the great capacity of the new electronic "giant brains" be harnessed to perform mental tasks beyond the mere rote?

The first computers seemed to be locomotives of thought. They were as big as locomotives, and performed feats equally awesome, outpulling hundreds of thousands of straining mathematicians on arithmetical calculations. Around 1955 a handful of enthusiastic young academics took up the challenge of extending their repertoire. The field was named *Artificial Intelligence* by John McCarthy at their first conference, at Dartmouth College in 1956. At the Dartmouth conference, Allen Newell, Herbert Simon, and George Shaw showed off their "Logic Theorist," the first working program of the new field. Starting with the axioms of number theory in Russell and Whitehead's famous (and mind-numbing) tome *Principia Mathematica,* the Logic Theorist was able to apply the rules of deductive inference to prove many of the theorems. Disturbingly, while the JOHNNIAC computer, on which Logic Theorist resided, could superhumanly do ten thousand numerical calculations per second, the program overall proved theorems no faster or better than a college freshman recently taught the subject. The program encoded symbols as numbers, needed hundreds of calculations to apply one rule of inference to one logical sentence, and explored many long chains of inference before finding a proof.

Programs were written during the next decade that proved theorems in geometry, solved calculus problems, and played good games of checkers and chess. Despite a number of clever mathematical and programming innovations, these programs remained at freshman competence. It also became apparent that each program could not be expanded beyond its initial narrow expertise. In particular, there seemed no way to give the programs "common sense" about everyday matters.

The Hard Easy

The discrepancy between the giant brain and the mental midget image of computers became worse in the late 1960s and early 1970s. Marvin Minsky's research group at the Massachusetts Institute of Technology and John McCarthy's at Stanford University connected television cameras and robot arms to their computers so "thinking" programs could begin to collect information directly from the real world. The early results were like a cold shower. While the pure reasoning programs did their jobs about as well and fast as college freshmen, the best robot-control programs, besides being more difficult to write, took hours to find and pick up a few blocks on a tabletop, and often failed completely, performing much worse than a six-month-old child. The disparity between programs that calculate, programs that reason, and programs that interact with the physical world holds to this day. All three have improved over the decades, buoyed by a more than millionfold increase in computer power in the fifty years since the war, but robots are still put to shame by the behavioral competence of infants or small animals. Computers have played grandmaster chess since the 1980s, and IBM's "Deep Blue" machine defeated Garry Kasparov, probably the best human player ever, in a May 1997 match. But Deep Blue needed a human assistant to see and physically manipulate the pieces. No robot existing could have done it in the wide range of circumstances Kasparov, or any child, finds trivial.

For machines, calculating is much easier than reasoning, and reasoning much easier than perceiving and acting. Why is this order of difficulty opposite for humans? For a billion years every one of our ancestors achieved that status by winning a competition for the essentials of life in a hostile world, often by a lifetime of sensing and moving more effectively than the competition. That escalating Darwinian elimination tournament bequeathed us brains spectacularly organized for perception and action—an excellence often overlooked because it is now so commonplace. On the other hand, deep rational thought, as in chess, is a newly acquired ability, perhaps less than one hundred thousand years old. The parts of our brain devoted to it are not so well organized, and, in an absolute sense, we're not very good at it—but, until computers arrived, we had no competition to show us up. Arithmetical calculation lies even further down the spectrum of human proficiency: it became a needed skill—for specialists—only in recent millennia, as civilizations accumulated large populations and pools of property. It is a tribute to our general-purpose perceptual, manipulative, and language skills, and our luck in finding compatible representations of numbers, that we can do arithmetic at all.

By a calculation later in this chapter, it takes a million times the power of a mid-1990s home computer to match a human brain doing what it does best—perception or motor control. Arithmetic lies at the other extreme of human performance. Occasionally a human is found who, through some biological quirk, can perform prodigious feats of mental calculation: some "lightning calculators" can add dozen-digit numbers as fast as they are presented and multiply them in a minute. Yet a home computer can best this performance many millionfold. Between these extremes, in certain tasks of rational thought, like playing chess, proving theorems, or diagnosing infections from lists of symptoms, present computers approximately match human performance.

When the reasoning problem strays beyond narrow logical confines, requiring some kind of visualization, or broad general knowledge, computers again lose out. To this day AI programs exhibit no shred of common sense—a medical diagnosis program, for instance, may prescribe an antibiotic when presented with a description of a broken bicycle because it lacks a model of people, diseases, or bicycles. Yet these programs, on existing computers, would be overwhelmed were they to be bloated with the details of everyday life, since each new fact can interact with the others in an astronomical "combinatorial explosion." [A project begun in 1985 called *Cyc* (from En*cyc*lopedia) at the Microelectronics and Computer Consortium in Austin, Texas, attempted to build just such a common-sense data base. Its founders estimated the final result would contain over one hundred million logic sentences about everyday things and interactions. The project continues today, but with the more modest claim of producing large databases for applications such as image retrieval.]

Machines have a lot of catching up to do. But catching up they are—for most of the century, machine calculation has improved a thousandfold every twenty years. The rate is twice as fast now, and developments already in the research pipeline should sustain it for at least several decades more. In less than fifty years, inexpensive computers will match and exceed—in raw information-processing power—even the well-developed functions of the human brain. But what are the prospects for programming these powerful machines to perceive, intuit, and think like humans?

Cockroach Race

Cybernetics attempted to copy nervous system function by imitating its physical structure. The approach stumbled in the 1960s on the difficulty of constructing all but the simplest artificial nervous systems, then regained vigor in the 1980s under the rubric *neural nets,* as computers became powerful enough to simulate interesting assemblies of neurons. Neural nets are taught to produce desired outputs from given inputs by tweaking the strength of interconnecting links in repeated training steps. In some pattern recognition and motor control tasks where the input/output relation is poorly understood but not too complex, they sometimes provide better performance for less effort than other programming techniques. However, with a maximum of a few thousand neurons, less than an insect's, present simulations cannot orchestrate elaborate behavior.

SCIENCE/COMPUTERS

"*Robot* is the most awesome work of controlled imagination I have ever encountered: Hans Moravec stretched my mind until it hit the stops." —Sir Arthur C. Clarke

From robotics pioneer Hans Moravec comes this mind-bending look at the future of intelligent machines and their genesis as Earth's next dominant race. Here, Moravec predicts that machines will attain human levels of intelligence by the year 2040, and that by 2050, they will surpass us. But even though Moravec predicts the end of human domination, his is not a bleak vision. Far from railing against a future in which machines rule the world, he embraces it, taking the startling view that intelligent robots will actually be our evolutionary heirs. "Intelligent machines, which will grow from us, learn our skills, and share our goals and values, can be viewed as children of our minds." And since they are our children, we will want them to outdistance us. In fact, in a bid for immortality, many of our descendants will choose to transform into "ex humans," as they upload themselves into advanced computers.

This provocative new book, the highly anticipated follow-up to his bestselling volume *Mind Children*, charts the trajectory of robotics in breathtaking detail. A must read for artificial intelligence, technology, and computer enthusiasts, Moravec's freewheeling but informed speculations present a future far different than we ever dared imagine.

"*Robot* is a dramatic, awe-inspiring prophecy of the human future by Hans Moravec. . . . His new book amplifies and substantiates that vision in concise, simple, yet elegant prose. *Robot* is an uncompromisingly radical synthesis of sociobiology, computer science, and philosophy. *Robot* paints a headbending but persuasive picture of our next 50 years, augmented with fascinating fragments from the more distant future." —*Wired*

"Intellectually adventurous and free with confident speculation. This is good knockabout stuff, a heady and unnerving glimpse into a possible future."

—Colin McGinn, *New York Times Book Review*

Hans Moravec, one of the leaders of robotics research, was a founder of the world's largest robotics program at Carnegie Mellon University. He is the author of *Mind Children: The Future of Robot and Human Intelligence.*

Cover design by David Stevenson

Cover art: Left, Robot Thinker *© Michael Knepper/SIS.*
Right, "Le Penseur" by Auguste Rodin, courtesy of Art Resource

OXFORD
UNIVERSITY PRESS
www.oup.com

90000
9 780195 136302
ISBN 0-19-513630-6