This Could Be Important

This Could Be Important

My Life and Times with the Artificial Intelligentsia

Pamela McCorduck

Carnegie Mellon University: ETC Press: Signature

Pittsburgh, PA

ISBN: 978-0-359-90133-3 (Print)
ISBN: 978-0-359-90134-0 (Hard Cover)
ISBN: 978-0-359-90138-8 (Digital)

Cover design by Heidi Bartlett
Copyedited by Rebecca Huehls

Text Permissions for *This Could Be Important*. Pamela McCorduck

Lines from “Tolstoy and The Spider” © 2011 Jane Hirshfield, from *Come, Thief.* NewYork: Knopf, 2011. Used by permission of the author.

“The Niagara River,” © Kay Ryan, from *The Niagara River*. New York: Grove Press, 2005. Used with permission of the author.

Lines from “A Pied in Arkansas” © J. Chester Johnson, from *St. John’s Chapel & Selected Shorter Poems,*second edition. Haworth, New Jersey: St. Johann Press, 2010. Used with permission of the author.

Image Permissions for *This Could Be Important*.

Watson: IBM’s Watson beats past *Jeopardy!* winners Brad Rutter and Ken Jennings in 2011. VincentLTE [CC BY-SA 4.0 (https://creativecommons.org/licenses/by-sa/4.0)] https://commons.wikimedia.org/wiki/File:WatsonPour203.png

Herb Simon: Herb Simon plays chess with CMU faculty member Bill Chase in 1973. Neil Charness, a PhD student who worked with Simon on his chess experiments, films. Courtesy of the Carnegie Mellon University Archives.

Herb Simon: Herb Simon lectures in Hamburg Germany in June 1977. Courtesy of the Carnegie Mellon University Archives.

Allen Newell: Allen Newell sitting at a prototype computer with a CRT monitor, designed to provide visual feedback to the user, 1975. Courtesy of the Carnegie Mellon University Archives.

Allen Newell: Allen Newell teaches a seminar course in 1977. Courtesy of the Carnegie Mellon University Archives.

John McCarthy: Copyright (c) The Board of Trustees of the Leland Stanford Junior University. All rights reserved. Courtesy of Stanford University Archives and Special Collections.

Marvin Minsky: Marvin Minsky and the "Minsky Arm" at the MIT Artificial Intelligence Lab. Ivan Massar, photographer; courtesy MIT Museum.

Edward Feigenbaum: Edward Feigenbaum (center), Director of the Stanford University Computation Center. 1966. Copyright (c) The Board of Trustees of the Leland Stanford Junior University. All rights reserved. Courtesy of Stanford University Archives and Special Collections.

Raj Reddy: Raj Reddy meets with students at Carnegie Mellon's Graduate School of Industrial Administration (GSIA) in 1989. Courtesy of the Carnegie Mellon University Archives.

Joseph Weizenbaum: Joel Moses (L) and Joseph Weizenbaum (R). Ivan Massar, photographer; courtesy MIT Museum.

Maja Mataric: Courtesy of Maja Mataric.

Ashley Montagu: Pamela McCorduck and Ashley Montagu at the opening of Santa Clara University's Technology Institute in January 1986. Image provided by Archives & Special Collections, University Library, Santa Clara University.

Lofti Zadeh: By Eastdept – Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=48614538

Computer Bowl: Pamela McCorduck's 1991 East Coast MVP trading card from the Computer Museum of Boston's yearly Computer Bowl competition. Copyright (c) The Computer History Museum. Used by permission.

Steve Jobs: Apple CEO Steven P. Jobs, left and President John Sculley present the new Macintosh Desktop Computer in January 1984 at a shareholder meeting in Cupertino, California, USA. AP Photo.

Harold Cohen: Harold Cohen in his studio. Photographer: Becky Cohen

Patrick Winston: Patrick Winston at his desk. Photo courtesy of Jason Dorfman, MIT CSAIL.

Elizabeth Honig: Elizabeth Honig addresses a conference in 2014 about her work. Photo: Jess Bailey.

Daniel Dennett: Photo Credit Irina Rozovsky.

Jeanette Wing: Jeanette Wing, 2010. Courtesy of the Carnegie Mellon University School of Computer Science.

Mary Shaw: Mary Shaw, 1976. Courtesy of the Carnegie Mellon University Archives.

Kai-Fu Lee: Portrait, Courtesy of Carnegie Mellon University School of Computer Science.

In memory of Joseph Traub, dearest of husbands for nearly half a century

For Edward Feigenbaum, dearest of friends for even longer

unless there is
a new mind there cannot be a new
line, the old will go on
repeating itself with recurring
deadliness.

—William Carlos Williams, Paterson

"Only puny secrets need protection. Big discoveries are protected by public incredulity."

—Marshall McLuhan

"I don't know whether I succeed in expressing myself, but I know that nothing else expresses me."

—Henry James, The Portrait of a Lady

Contents

PART ONE: THE TWO CULTURES

PART TWO: BRAINS

PART THREE: CULTURE CLASH

PART FOUR: THE WORLD DISCOVERS ARTIFICIAL INTELLIGENCE

PART FIVE: SILICON VALLEY SKETCHBOOK

PART SIX: ARTS AND LETTERS

PART SEVEN: AND WHEREFORE WAS IT GLORIOUS?

Part One: The Two Cultures

"I believe the intellectual life of the whole of Western society is increasingly being split into two polar groups…Literary intellectuals at one pole…at the other, scientists…Between the two, a gulf of mutual incomprehension—sometimes (particularly among the young) hostility and dislike, but most of all lack of understanding."

—C.P. Snow, *The Two Cultures,* 1959

Living in the Exponential

1.

Stanford University's climate is so benign that one luminous day dissolves into another, almost without distinction. On a radiant day in 1965—spring? summer? autumn?—a world-famous but jet-lagged and weary Soviet computer scientist visited the Stanford computer science department. He was Andrei Yershov, and I was asked show him around.

He particularly wanted to see something called the Doctor program in action, one of the first-ever interactive computer programs. Doctor's home was the Stanford Artificial Intelligence Laboratory, a few miles away, but graduate students had rigged a remote hook-up on the second floor of Polya Hall on the main Stanford campus, which terminated in a teletype machine. This was an ordinary keyboard on a stand, used to transmit and receive text messages. You typed your message to transmit it; in receiving, the keyboard depressed, letter by letter, with ghostly autonomy. It had no screen: the exchange was recorded by physical strikes on a continuous roll of paper.

Yershov sat down at the teletype, and the graduate students and I gathered around him to watch.

A loud electric hum suddenly filled the room; the teletype clattered to life, and banged out a message:

> DOCTOR: This is the doctor. Good afternoon. Please type your name.

The visitor did, and the Doctor program warmed up with some polite preliminaries. Then:

> DOCTOR: What brought you here to see me today?
>
> VISITOR: Oh, nothing much. I'm feeling a bit tired, that's all.
>
> DOCTOR: Why do you think you're feeling a bit tired?
>
> VISITOR: Well, I've been traveling a lot, and away from home.
>
> DOCTOR: Tell me about your family.

The exchange suddenly turned intimate. Our visitor knew perfectly well that a computer, not a human being, was at the other end. Nevertheless, he began to disclose his worries about his wife, his children, his distance from them—both geographical and emotional. We watched in painful embarrassment, trying hard not to stare, yet mesmerized. Something about that impersonal machine had evoked a response from the visitor forbidden by the norms of polite human conduct. If a sophisticated computer scientist could succumb to such a revealing conversation, insensible to the spectators over his shoulder, what would a less sophisticated person make of it?

By now I'd been around the field of artificial intelligence some five years, since 1960. But that moment, something changed for me. This wasn't chess and checkers; this wasn't proving theorems or solving puzzles. It wasn't any of the other abstract tasks to which AI had been applied. This was a connection between two minds, one human, one —Other. It was uncanny. Unreal. Yet altogether real. Explicable. Yet not.

Some epiphanies are evanescent. They burst upon you, then hide for years. So it was here. This epiphany would come back every so often in its revelation of a machine's mind. I never thought it was wicked. I didn't think it was inappropriate. But I thought about it. I'd think about it the rest of my life.

On that sunny afternoon, I couldn‘t know that, one way or another, AI would preoccupy me for decades. That I was poised for a journey, leading from a place of friendly curiosity all the way to conviction, sometimes moving past cartoon versions of the field (and of me), ducking best-selling jeremiads and scathing mockery about it all. At times the path seemed to lead me through the looking glass to an odd and disquieting world, a place that turned everything I thought I knew upside down, inside out. But I'd emerge at last to a place of relatively serene and optimistic understanding, along with some grave misgivings.

Each of us will take this journey in the future. This book is an invitation to follow me. Maybe the path is easier, knowing someone has stumbled along it before. And I urge readers impatient with the technical details to skip ahead to the next place where the personal emerges again.

Though much sound and fury has swirled around whether machines *really* think (the way humans do) or are only faking it, that tired dispute bores me. An old behaviorist trope holds that although birds and airplanes don't operate on exactly the same principles, both birds and airplanes fly. Actually, the Bernoulli principle underlies them both. I feel the same way about thinking brains and thinking computers. Unless you're a cognitive scientist, devoted specifically to modeling human cognitive behavior, why get exercised about the authenticity of thinking? Because, disputants say, it's not that human cognition marks our superiority as a species—though the passions, and fury, which people bring to the topic make me suspect that this is what they're defending. No, they say, if the machines don't think like us, then how can they share our values? That could be. But many humans wouldn't recognize "our" values either. I'm not sure I recognize yours.

2.

Nearly half a century after that Stanford afternoon, on Valentine's Day 2011, and for two distinctly cold nights following, I watched the old gladiators of human and machine intelligence clash again. Their coliseum was the TV quiz show *Jeopardy!* On the show, contestants need to quickly access lots of trivial knowledge; figure out riddles, puns, and jokes; and interpret ambiguous statements. Representing human intelligence were the two best players *ever* of this game. The two humans stood at podiums, and between them was a big blue (of course) box, the avatar of Watson, a computer program designed by a team at IBM. Its logo reminded me of Keith Haring's "Radiant Baby."

Who would triumph, man or machine?

As I watched, that 1965 afternoon at Stanford came back, Yershov and the wheezing, clattering teletype. How far things had come. A small host of benign spirits crowded into my Manhattan living room, men I'd known over half a century, whose work helped to bring all this about: Herb Simon, Allen Newell, John McCarthy, and

Marvin Minsky. They were founding fathers of artificial intelligence, American geniuses all of them.

The first evening, responding in a not-quite-human voice,[1] Watson barely pulled ahead of its human competition. The second evening, the humans might've come back, but no, this time Watson-the-machine pulverized its human competition. The third evening was a mop-up—Watson grabbed three points for every one point of its nearest human competitor. The spirits surrounding me smiled, and so did I.

AI's accomplishments were suddenly public and heady: now came nearly accident-free, self-driving cars (and their hacker perils); phones activated by recognizing your face and responding to your voiced instructions; applications that began to transform entire fields: law, finance, medicine, science, engineering, entertainment. And yes, spying. Daily, the applications continue to cascade out.

A grand spiral nebula of the sciences, statistics, mathematics, logic, and dazzling engineering has swirled together to create modern artificial intelligence.

In the next few years, IBM's Watson, for example, marched on: medical research and clinical applications, business and financial applications. Watson was poised to answer questions you hadn't yet asked. In 2015, Watson held the red carpet at the Tribeca Film Festival in New York City, ready to be your partner as screenwriter or designer. In 2017, the program did a few weeks as an art guide

[1] Technologists then and now have debated how close to "human" a synthesized voice should—or might—sound. See John Markoff (2016, February 15). An artificial, likable voice. The New York Times. Further disputes have arisen about the gender of such voices: do we want a stereotyped subservient female voice, or a male voice mansplaining?

in the Pinacoteca of São Paulo, Brazil, answering direct questions from visitors about individual paintings. No one would guess what a fraught relationship IBM once had with AI, but I remember, and smiled again. Google, moving in other directions, promotes the self-driving car, machine-reads and digests tons of text and images; in London, DeepMind, another subsidiary of Google's parent company, Alphabet, has produced both the chess champion and the Go champion of the world, a feat long considered impossible. DeepMind's program, first called AlphaGo, and then AlphaZero, because it started knowing nothing but the rules of the game, developed its skills by playing against itself, not by taking instruction from humans. Therefore it could be said to understand the game, and unlike previous game playing programs, didn't use brute force. With more possible positions in a Go game than atoms in the universe, according to a January 2016 blog post by Demis Hassabis, the co-founder and CEO of DeepMind, the AlphaZero program will lead the way to general, as opposed to specialized, artificial intelligence. Maybe.

About AlphaZero as chess champion, mathematician Stephen Strogatz wrote: "Most unnerving was that AlphaZero seemed to express insight. It played like no computer ever has, intuitively and beautifully, with a romantic, attacking style. It played gambits and took risks....Grandmasters had never seen anything like it. AlphaZero had the finesse of a virtuoso and the power of a machine. It was humankind's first glimpse of an awesome new kind of intelligence." (Strogatz, 2018). James Somers sensitively describes the reactions to AlphaZero of the human champions: first sadness, depression, at last acceptance. "The algorithm behind AlphaZero can be generalized to any two-person, zero-sum game of perfect information (that is, a game in which no hidden elements exist, such as face-down cards in poker)…..At its core was an algorithm so powerful that you could give it the rules of humanity's richest and most studied games and, later that day, it would become the best player there has ever been. Perhaps more surprising, this iteration of the system was also by far the simplest" (Somers, 2018). AlphaZero has since made inroads on multi-person games.

As the second decade of the 21st century came to an end, artificial intelligence was on every editor's uneasy mind, and the phrase springs up daily in broadcasts, newspapers, journals, and blogs. Within one week in May 2018, Carnegie Mellon announced the creation of the first undergraduate degree with a major in artificial intelligence[2]; the White House held a summit meeting on the future of AI with representatives of major corporations[3]; *The New York Times* quoted Sundar Pichai, Google's chief executive, who said that advancements in artificial intelligence had pushed Google to be more reflective about its corporate responsibilities around AI[4]; and *The New Yorker* ran an article entitled "Superior intelligence: Do the perils of A.I. exceed its promise?"[5] In 2019 *The New York Times* presented an entire section on the ethics of AI.[6]

Suddenly, the idea of the Singularity—that strangely Oedipal

[2] In October 2018, MIT topped CMU's establishment of a mere major in AI by announcing an entire School of Artificial Intelligence to be supported by a billion-dollar commitment.

[3] Shepardson, David. (2018, May 10). White House to hold artificial intelligence meeting with companies. Reuters.

[4] Wakabayashi, Daisuke. (2018, May 9). Google promotes A.I. but acknowledges tchnology's perils. The New York Times.

[5] Friend, Tad. (2018, May 14). Superior Intelligence: Do the perils of A.I. exceed its promise? The New Yorker.

[6] The New York Times, "Artificial Intelligence: Ethical AI." March 4, 2019.

crossroad when machines become smarter than their creators—began to appear like maybe more than science fiction.[7]

Larry Birnbaum, a professor of computer science at Northwestern, calls it “living in the exponential.” An exponential curve seems to move gradually at first. Then it begins to climb more steeply. And climb ever more steeply. We all now *live in the exponential* of artificial intelligence.

3.

For sixty years, I’ve lived in AI’s exponential. I’ve watched computers evolve from plodding sorcerer’s apprentices to machines that can best any human at checkers, then chess, then the guessing game, *Jeopardy!*, and now the deeply complex game of Go. By 2018, about two-thirds of American adults had adopted AI into their pockets and handbags, in the form of smartphones. Although only 15% of Amercian households had voice-activated smart speakers such as Alexa or Echo, their adoption rate was outpacing smartphones and tablets.

More technically, according to the 2017 AI Index, AI became better at tasks that once required human intelligence. AI’s error rate in labeling images declined from 28% in 2010 to less than 3% in 2016. (Human error at this task is about 5%.) In 2017, speech recognition

[7] Science fiction writer Vernor Vinge has often been cited as the originator of this phrase to describe such a moment in the evolution of intelligence, but John von Neumann or Stanislaw Ulam seem to have been the first to appropriate the phrase from mathematics. In his memorial tribute to von Neumann, Ulam writes: “One conversation [we had] centered on the ever accelerating progress of technology and changes in the mode of human life, which gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue.” Ulam, Stanislaw. (May 1958). John von Neumann, 1903-1957. Bulletin of the American Mathematical Society, 64, 3. My thanks to Bruce Garetz for bringing this to my attention.

achieved parity with humans in the limited domain of Switchboard, an app that looks up people's contact information and connects them. Two programs, one from Carnegie Mellon and one from the University of Alberta, beat professional experts in poker, and a Microsoft program achieved the maximum of points in Ms. Pac-Man on an Atari 2600. However, the index also acknowledges that these champion programs falter if the task is altered even slightly.

At first a friendly skeptic, I slowly came to believe that AI is inevitable. To push beyond human limitations has been our collective obsession for more than half a millennium, since the beginning of the Scientific Revolution. Maybe longer. At the dawn of that revolution, we learned that knowledge is power—power in the sense that Francis Bacon meant when he coined the phrase: power to shape our environment, our health, our fortunes, perhaps our future. AI is now a fundamental part of that.

Despite its dubious early reputation, AI began at once to elucidate the nature of human intelligence and pushed scientists to look at intelligence in other species. As it matured, the field began to propose principles of intelligence that might govern both the biological and social worlds. Certainly thanks to AI, the very horizons that human intelligence yearns toward have dramatically expanded. AI and its sibling cognitive sciences suggest that if we're pre-Newtonian in our understanding of the laws of intelligence, perhaps such laws exist and can be discovered.

Meanwhile, AI is transforming everything, including what I studied in college, the humanities. Up to now, assertions or hand-waving have mostly defined key questions in the Western tradition—what is thinking, memory, self, beauty, love, ethics? But in AI, the questions must be specified precisely, realized in executable computer code. Thus, eternal questions are re-examined. Some people think this somehow diminishes human achievement in the arts, the humanities, and philosophy. Even near-centenarian Henry Kissinger declared that AI marks the end of the Enlightenment—a statement to give pause for many reasons (2018). But I believe that AI helps us understand and perhaps answer those vital questions better.

For several decades, people like Ray Kurzweil, an engineer, inventor, and futurist, have talked about AI's imminent, inevitable, full, and lush arrival in the form of The Singularity. In a 2012 PBS NewsHour interview with Paul Solman, Kurzweil said, "Artificial intelligence will reach human levels by around 2029. Follow that out further to, say 2045, we will have multiplied the intelligence, the human biological machine intelligence of our civilization a billion-fold." He's a professional futurist. It's his job to be provocative. Although sometimes he admits he might be off by a few decades, he assures us the eventual result will be the same. Exceptions to this view, however, are plenty, as we'll see. [8]

More than forty years ago, Ed Fredkin, then at MIT, said he expected that, to mature AIs, humans would be dull—maybe they'd keep us as pets. This possibility played out in Spike Jonze's 2013 movie, *Her*, where the AIs didn't revolt against humans or try to conquer them as they had in so many classic tales. Instead, the bored AIs abandoned humans (and left humans yearning for the return of their brainy pals).

[8] Many people strongly disagree with this scenario. For example, a 2016 panel of AI experts convened to examine the ethical and instrumental consequences of AI disagreed. Kevin Kelly, the emeritus editor-in-chief of Wired and a perceptive observer of technology for four decades, disagrees. John Markoff, who covered Silicon Valley for The New York Times for many years, also disagrees. I'll take this up later.

Most philosophers, whose response to AI over its early decades was was a species of mind games, parables and fables to prove the technology wasn't *really* thinking, began at last to pay serious attention. In 2010, David Chalmers, an NYU philosopher, presented a set of reasonable scenarios to address the Singularity and invited colleagues from all over the world to respond.[9] Nick Bostrom, a philosopher and cognitive scientist at Oxford University, sees AI's possibilities and dangers and concludes that meeting its challenges successfully, especially learning to control it, is the essential human task of our century (2014). In *From Bacteria to Bach and Back: The Evolution of Minds,* philosopher Daniel Dennett, long a friendly observer and critic of artificial intelligence and a deep thinker about matters of intelligence generally, offers a wonderfully nuanced view of the whole situation because he sees issues of intelligence in context and as parts of a great whole (2017).[10]

Predicting the future of AI isn't just for scholars. In an interview with Julian Sancton, Beau Willimon, creator, developer, and producer of the television hit *House of Cards*, says:

> This is where I sound like a complete fucking madman. But I actually think that humans are only the beginning. I think we've moved beyond biological evolution, which is incredibly inefficient. It took 15 billion years for us to get where we are. And there's still 15 billion more years to go. We are the salamanders crawling out of an ocean. And there's something way beyond. And I don't think it's God that created the universe. I think it's the universe's project to create God. And the things that we do in our rudimentary ways are teaching the next thing how to

[9] Chalmers' article appeared in Journal of Consciousness Studies and was followed by responses from many in another 2012 issue of Journal of Consciousness. See "The Singularity: Ongoing Debate Part II."19, (7-8).

[10] Dennett and I nearly always end up in the same place, but he does the careful, heavy thinking. I just barge ahead, grateful to outsource that careful, heavy thinking to him.

> imagine. And then it's going to take it from there. It will ask questions we don't even know how to ask. It will think the things we are incapable of thinking. It will experience and feel the things that we aren't capable of. (Sancton 2014)

It's certainly a point of view.

It will ask questions we don't even know how to ask. It will think the things we are incapable of thinking. It will experience and feel the things that we aren't capable of. Yes, I believe that will happen eventually. We think of AI in terms of personal gadgets—my search engine will be better, my car will drive itself, my doctor will be better able to heal me, my grandma can be safely left home alone as she ages, a robot will finally do the housework. But greater contributions of AI will be planetary, teasing out how the environment and human wellbeing are subtly intertwined. Al's greatest contribution might be its fundamental role in understanding and illuminating the laws of intelligence, wherever it manifests itself, in organisms or machines.

For a long time, I've been comfortable with such ideas. Unduly optimistic, maybe, but I look forward to having other, smarter minds around. (I've always had such minds around in the form of other humans.) I don't much worry they'll want my niche—though that presupposes a planet that won't, in one of Bostrom's scenarios, be tiled over entirely with solar panels to supply power for the reigning AI. Humans will endure, but possibly not as the dominant species. Instead, that position might belong someday to our non-biological descendants. But really, the scary future scenarios sound as if humans have no agency here. We certainly do, and as we'll see, it's already at work.

A search (powered by AI techniques, of course) will quickly show

how we've already woven AI around and inside our lives, turning scientific inquiry into human desire, even stark necessity. When we did not—the nuclear catastrophe of Fukushima Daiichi, for example—we wished we had. AIs fly, crawl, inhabit our personal devices, connect us with each other willingly or not, shape our entertainment, and vacuum the living room.

Robots, a particularly visible form of AI (*embodied,* in the field's term), occupy a significant space in our imaginations, their very birthplace. Books, movies, TV, and video games provoke us to conjecture about some of the ways we might behave and the issues embodied AIs will raise when they become our companions. But this visible embodiment, humanoid or otherwise, is only one form AIs will take. The disembodied, more abstract intelligences, like Google Brain, AlphaZero, and Nell[11] at Carnegie Mellon are hidden inside machines invisible to the human eye, scoffing at human boundaries. Their implications are even more profound.

Distributed intelligence and multiagent software inhabit electronic systems all over the globe, seizing information that can be studied, analyzed, manipulated, redistributed, re-presented, exploited, above all, learned from. Human knowledge and decision-making are rapidly moving from books and craniums into executable computer code. But fair warnings and deep fears abound: algorithms that take big data at face value reinforce the bigotries those data already represent. Bad enough that data about you you're aware you're volunteering (submitted for drivers' licenses for example) are collected, aggregated, and marketed; much worse that involuntary data collected from your on-line behavior (your purchases, your use

[11] In this book, I will not capitalize most program acronyms or abbreviations, except for initial caps. It's unnecessary and tiring to the reader's eye.

of public transportation) is also a profit center and a spy on you. Horrors have crawled up from the dark side: bots that lie and mislead across social media, trolls without conscience, and applications whose unforeseeable consequences could be catastrophic.

Larry Smarr, founding director of the California Institute for Telecommunications and Information Technology, on the campus of the University of California at San Diego, calls this distributed intelligence and multiagent software the equivalent of a global computer. “People just do not understand how different a global AI will be with all the data about all the people being fed in real time,” he emailed me a few years ago. By sharing data, he continued, the whole world is helping to create AI at top speed, instead of a few Lisp programmers working at it piecemeal. The next years will see profound changes. In short, AI already surrounds us. *Is* us.

The industrialization of reading, understanding, and question-answering is well underway to be delivered to your personal device. Some of these machines learn statistically; others learn at multiple, asynchronous levels, which resembles human learning. They don’t wait around for programmers but are teaching themselves. Understanding the importance of this, many conventional firms like Toyota or General Electric are reinventing themselves as software firms with AI prominent.

Word- and text-understanding programs particularly interest me, partly because I’m a word and text person myself, and partly because words, spoken or written, at the level of complexity humans do them, seem to be one of the few faculties that separate human intelligence from the intelligence of other animals. (Making images is another.) Other animals communicate with each other, of course. But if their communication is deeply symbolic, that symbolism has so far evaded us. Moreover, humans have means to communicate not only face to face, but also across generations and distances, and we do so orally, then by pictorial representations, by speaking, creating pictures, writing, print, and now by electronic texts and videos.

For a long time, we were the only symbol-manipulating creatures on the planet. Now, with smarter and smarter computers, we at last have symbol-manipulating companions. A great conversation has begun that won't be completed for a long time to come.

What follows is one story (there are many) of an extraordinary half-century and more, when humans edged toward an epochal event: a new kind of intelligence emerged, designed by us, to live beside our own. But this book is about humans, not machines. As it happens, AI's coming of age, if not its full maturity, has paralleled my own life. So this is the saga of a grand scientific quest, intertwined with my personal quest. It's a coming of age story of a scientific field and of a naïve young woman—now slightly wiser and decidedly older.

I kept a journal—I still do—for I sensed what I witnessed would be momentous.[12] So much was to happen as a consequence of AI. For good or otherwise, everything—social life, medicine, transportation, communication—has changed. But for a long time, I'd spend my life pulling on the sleeves of serious thinkers, trying to tell them that this—artificial intelligence—could be important.

I offer a personal story because it's the particulars that illuminate: personalities, friendships, enmities, context, chance. To grasp these early times, abstractions won't do.

[12] The original hand-written spiral-bound notebooks are archived in Hunt Library at Carnegie Mellon University.

The Capacious Structure of Computational Rationality, Fast and Slow Thinking, an Intelligence Continuum

1.

In my lifetime, the foundations have been laid for a capacious and elegant structure whose completion will take many generations. Like a medieval cathedral, or better, the great Hagia Sophia, it will be a new temple of holy wisdom. The structure will subsume the many instances of intelligence wherever it's found—in brains, minds, or machines; in cells, trees, or ecosystems—under general principles, perhaps even laws. Already this structure has begun to shelter and illuminate new definitions of intelligence as it's slowly and meticulously formed from observation, experimentation, modeling, and example. It may even finally produce an authentic metric for intelligence, which, more than a century after intelligence testing began, still eludes us.

This ambitious new effort aims to discover the laws of intelligence the way Newton discovered the laws of motion. Before Newton, no one quite saw the commonalities among a stroll in the park, the turbulence of a river, the winds, the tides, the circulation of blood, the rolling of a carriage wheel, the trajectory of a cannon ball, or the paths of the planets. Then Newton found the underlying generalties that, at a fundamental level, explained and connected them (and so much more). Varieties of intelligence may be even more abundant than varieties of motion, but the fundamental laws that underlie them, should they be found, will be simple and will elegantly subsume that infinite variety.

Computer scientists have begun to call this edifice *computational rationality*, a converging paradigm for every kind of intelligence (Gershman et al., 2015). The structure is inspired by the general agreement that intelligence arises not from the medium that embodies it—whether biological or electronic—*but the way interactions among elements in the system are arranged.* Intelligence begins when a system identifies a goal (I want to go to the movies; I need to learn analytic geometry), learns (from a teacher, a training set, its own experience or that of others), and then moves on autonomously, adapting to a complex, changing environment.[1] Or you might imagine intelligent entities as networks, often arranged as hierarchies of intelligent systems—humans certainly among the most complex, but congeries of humans even more so.

Three core ideas characterize intelligence. First, intelligent agents have goals, form beliefs, and plan actions that will best reach those goals. Second, calculating ideal best choices may be intractable for real-world problems, but rational algorithms can come close enough

[1] Cognitive, computer, and neuroscience work closely with each other, but AI, a branch of computer science, is the only field that attempts to build machines that will function autonomously in complex, changing environments. As a consequence, AI has made rigorous the study of intelligence wherever it appears. In the earliest AI research, this overarching paradigm of the nature of intelligence was implicit, but not conspicuously self-evident.

(*satisfice* is Herbert Simon's term) and optimize the costs of computation. Third, these algorithms can be rationally adapted to the organism's specific needs, either off-line through engineering or evolutionary design, or online through metareasoning mechanisms that select the best strategy on the spot for a given situation (Gershman et al., 2015).

Our unfinished—our barely begun—grand structure of computational rationality is already large and embraces multitudes. For example, biologists now talk easily about cognition, from the cellular to the symbolic level. Neuroscientists can identify computational strategies shared by both humans and animals. Dendrologists can show that trees communicate with each other to warn of nearby enemies, like bark beetles ("Activate the toxins, neighbor!") or admonish the children ("Not so fast, sapling").

The humanities are comfortably at home in this structure, too, although it's taken many years for most of us to see that. And of course here belongs artificial intelligence, a key illuminator, inspiration, and provocateur.

To grasp this fully, we must begin by abandoning old beliefs. One held that only humans could embody *real* intelligence (or *strong intelligence* in the phrase of philosopher John Searle). *Artificial* intelligence, no matter what it achieved, was different, and therefore lesser—*weak intelligence.*[2]

[2] At The AI Summit conference in 2014, leading AI researchers used these two phrases (with the same precision as the philosopher, which is to say, not much at all). At first I took it as irony. Then I thought they'd seized these terms the way that gay people reclaimed "queer" as in-your-face defiance of critics. No, my ahistoric friends had no idea where the phrases had come from, found them useful, and employed them innocently. When I told this to philosopher Daniel Dennett, who's had some spirited public exchanges with John Searle, he just groaned. But apparently the phrases are here

We must also let go of the old belief that intelligence resides solely in an individual cranium. This is a hard re-set for anyone who's grown up in Western culture, which has traditionally emphasized individual intelligence over its collective nature.

Not surprisingly, some people object to the cognitive, the computer, and the neurosciences exploring and replicating the mind, a patch thought to be uniquely human. To this opposition, philosophers are the mind's sole interpreters. In the daily Columbia University *Spectator,* an undergraduate complains that he's okay with reading Kant and Hume on the nature of mind as required by the Core Curriculum, but why is the *science* of mind off in some science ghetto? Why can't he read the new findings about the mind alongside the speculations of Kant and Hume?

Why not indeed?

Fanciful maybe, but think of intelligence as a continuum. At one end of the continuum are simple cells figuring out what they need to survive to avoid self-destruction. At the other end are humans exhibiting wide-ranging if not entirely general-purpose intelligence over many different kinds of situations and manipulating symbols through storytelling and making images, in ways no other organisms seem to do.

A bacterium doesn't *think* about seeking energy from whatever source powers its metabolism in a thoughtful logical way. It just goes for those specialized energy bars in its surroundings. A cheetah doesn't think, "Yum, would that critter make a nice dinner? Should I invite those tedious people next door?" The cheetah automatically

to stay until they're better defined or revealed as nonsense. (I adopted the once derisive term "artificial intelligentsia" for the title of this book because it amused me.)

identifies (extracts the features of) prey and pursues it as fast as possible. We've called this mere instinct, but in fact it's a kind of intelligence at work, perceiving, recognizing, and quickly acting on those perceptions and recognitions.

Similarly, machine learning (ML) quickly perceives and recognizes patterns in large amounts of data. Specifically, machine learning is an array of algorithmic, statistical, and mathematical techniques that can improve automatically through experience. ML relies on enormous data sets to find patterns and explore nuances. Among its varieties are supervised learning, unsupervised learning, reinforcement learning, deep learning, and neural nets—these last two brain-like, but not brains. Thus ML could be said to correspond to the intelligence of certain organisms, from simple cells to complete animals, paralleling what we call instinct in such organisms.

As remarkable as they are, ML applications are narrow. They cannot move from domain to domain and fail if initial conditions change even slightly. Humans are still needed to label the initial patterns that the algorithms evoke (*cat*, *melanoma*, *paper shredder*, *road obstacle).* We now acknowledge that the algorithms and statistical methods ML uses are not neutral. Human beings with cultural biases, conscious and unconscious, construct them. Thus, commonplace assumptions (sometimes false) and deep human prejudices of the moment are baked into the novel patterns ML teases out of big data.

Yet in some simplified way, ML mimics some of the functions of the organic brain. MIT's Tomaso Poggio, an eminent researcher across neuroscience and computation, reminds us that the recently successful algorithms behind AlphaZero, now the global Go champion, and Mobileye, a vision-based collision-avoidance system for drivers, are based on two algorithms originally suggested by discoveries in neuroscience: deep learning and the associated techniques of representation, reinforcement learning, transfer learning and other techniques.

At the other end of the intelligence continuum is the kind of symbolic cognition that humans exhibit, which is slow and analytical (over seconds, minutes, even hours and days), abstract, logical, and heuristic (based on rules of thumb and knowledge other humans, such as teachers, or texts, or experience have provided). The part of AI that corresponds to that kind of human intelligence is sparsely populated with applications. Andrew Moore, former dean of the school of computer science at Carnegie Mellon calls AI "the science and engineering of making computers behave in ways that, until recently, we thought required human intelligence." *Until recently* changes over time.

Yes, humans exhibit both kinds of intelligence, what psychologist Daniel Kahneman calls thinking fast and thinking slow, because humans evolved from the end of the intelligence continuum we share with all organisms.[3] At the moment, ML applications (thinking fast) however narrow, abound, and will proliferate, it seems, forever. Symbolic cognition applications (thinking slow) are few and far between although artificial intelligence had its birth in those applications.

More than one AI researcher has recently told me that ML has pretty

[3] We now know human thinking is strangely and strongly affected by the composition of human gut flora and fauna, so my gastroenterologist and I are in lively discussions about whether machines, lacking guts, will ever be able to think like humans. Perhaps supplying machines with guts and the appropriate biome is the missing link to human-like thinking in machines. Perhaps that's a terrible idea. Thanks to Jonelle Patrick for raising the question to me in the first place.

much run its course in terms of *research.* “Breakthroughs” celebrated almost daily in the media are really new *applications* of ML, that however brilliant and useful, cannot move between domains, and it bears repeating that they fail if the initial conditions are even slightly changed. ML also elides the embarrassing and deeply consequential fact that researchers often cannot explain the inner workings of their mathematical models. Because they lack rigorous theoretical understanding of their tools, Ali Rahimi, a well-known machine-learning researcher, said deep-learning researchers are working like alchemists instead of scientists (Naughton 2018).[4] This is not to slight how these narrow applications can still have significant effects (especially see Chapter 30 on China and the United States.).

On the other hand, MIT’s Patrick Winston has said that symbolic cognition has particular characteristics. It can merge two expressions to make a larger expression without disturbing the two merged expressions. This aspect of symbolic cognition allows humans to build complex, highly nested symbolic descriptions of classes, properties, relations, actions, and events. Winston and his colleague Dylan Holmes (2018) write: “With that ability we can record that a hawk is a kind of bird, that hawks are fast, that a particular hawk is above a field, that the hawk is hunting, that a squirrel appears, and that John thinks the hawk will try to catch the squirrel.” Although other animals might have internal representations of some aspects of the world, they seem to lack complex, highly nested symbolic descriptions.

[4] I’m taken aback to hear AI related to alchemy once again, as it was in the 1960s. I’ll spare my readers the essay I could write about how science evolves. But Rahimi is correct that these mysterious applications are being used in real life right now without deep understanding of how they work. So was aspirin mysterious for years after it was deployed, but somehow AI does seem more momentous than aspirin.

So symbolic cognition may be the next *research* frontier, perhaps a return to Good Old Fashioned AI, or GOFAI as it's known, but brought up to date with improved technology (Somers, 2017). Or that new research frontier may be something altogether different.

The totality of intelligence is variegated, collective, distributed, even emergent. Understanding and knowledge are enacted only within a larger system. Nothing resides or is born solely in a single human's head.

The first inkling of this I got was from a young scientist (his name lost to me) sometime in the early 1970s. We were strolling beneath the eucalyptus trees on the Mills College campus in Oakland, and he was trying to explain the systems approach to intelligent behavior. He didn't use that phrase; he may not have known it. You and I think of ourselves as intelligent, he said, but we didn't invent the language we speak. No matter how brilliant we are, we didn't invent much of anything, compared to how much we rely on the inventions and innovations of countless others in the past and present.

I stopped. I knew at once he was right. For centuries, Western thought has been strongly biased toward celebrating the individual, his consciousness, creativity, insight, or brilliance without mentioning the milieu this consciousness, creativity, insight, and brilliance finds itself in, draws upon, and recombines to create novelty.

The assumption that intelligence is the property of the individual alone is so foundational in Western thought that hardly anyone thinks to question it, at least in the First Culture, whose literary tradition I learned. The Second Culture, science and mathematics, does a better job of balancing the credit between those who came before and the work of the individual by referring explicitly to a chain of precedents in the form of citations. Yes, some gifted individuals move it all forward, sometimes brilliantly. But they do so inside, relying upon a system that they alone didn't invent.[5]

As AI research fills out the nearly empty, slow-thinking end of the intelligence continuum, the symbolic part, I believe that the structure of computational rationality—the principles, the laws of intelligence—will at last be revealed.[6]

2.

A history exists of all this, a human story about the invention of artificial intelligence by a handful of brilliant scientists who understood that computers could exhibit what we call intelligence, if only they—scientists and machines—worked at it. At the time, the idea of artificial intelligence was audacious, a bit loony, and the stuff of science fiction, not science.

The earliest researchers were not all men. Margaret Masterman, a former student of Ludwig Wittgenstein, established the Cambridge Language Research Unit at Cambridge University in 1955 (though not officially a part of the university). The unit pursued automatic translation, computational linguistics, and even early quantum physics. Thus her efforts were contemporary with those of Allen Newell and Herbert Simon, who are generally credited with creating the first working AI program. Masterman's work and that of her associates had pioneering importance in machine translation, but linguistics and machine translation were soon parted from core AI.

[5] This is now explicit in books like Sloman, Steven, & Fernbach, Philip. (2017). The Knowledge Illusion: Why We Never Think Alone. New York: Riverhead.

[6] I'm grateful for discussions with Edward Feigenbaum to clarify my own intuitions about the intelligence continuum.

(Why language wasn't considered symbolic baffles me.) Until someone writes a seriously revised history of AI, correcting in some ways my own *Machines Who Think*, Masterman won't get the credit she deserves.[7]

So AI, as first imagined, a field that operated at levels of *symbolic* human intelligence, counts its founding fathers as all American men. This was much the result of post-World War II United States prosperity. Alan Turing, a brilliant Englishman, had certainly foreseen the possibilities of computer intelligence and even designed, though didn't program, a primitive chess-playing machine. He proposed "the imitation game," which famously came to be called the Turing test. A set of human judges must conduct a freewheeling conversation (in text), the kind of *viva voce* beloved by Oxford and Cambridge, with respondents who might or might not be computers concealed from the judges. By the human qualities these conversations exhibited, the judges were to decide whether respondents were computers or a humans.[8]

[7] Thanks again to Edward Feigenbaum for bringing Masterman to my attention.

[8] The Turing test lacks refinements but contests are held annually, with the rule that thirty percent or more of the judges must agree on the "humanness" of a respondent. In Summer 2014, for the first time, a third of the judges agreed that Eugene Goostman was a charming, maybe typical, 13-year-old Ukrainian boy, who liked hamburgers and candy and whose father was a gynecologist. However, Eugene was a program put together by a team led by Russian Vladimir Veselov and Ukrainian Eugene Demchenko. As professionals scoffed, two of the scientists who conducted this experiment wrote a long clarification for the Communications of the ACM (April 2015) regarding numbers of judges, judges' knowledge, and quoted Turing: "Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's?" Moshe Y. Vardi, editor-in-chief of the journal responded tartly: "The details of this 2014 Turing test experiment only reinforces my judgment that the Turing test says little about machine intelligence. The ability to generate a human-like dialogue is at best an extremely narrow slice of intelligence." Not so negligible, I'd say. As we'll see, in the next few years, human/machine conversation became much more sophisticated.

Turing was prevented from realizing what he was sure computers could do not only by British post-WW II national austerity, but also by British peevishness and factionalism. (Manchester? you can hear the London boffins say to each other, as they clutch scarce postwar British research funds. Manchester? *Really?*) Finally, Turing was hounded to a premature death by British laws that criminalized his homosexuality and drove him to suicide.[9] No amount of subsequent pardons and regrets can change this or compensate for the loss.

Perhaps even before Turing, certainly contemporary with him, the German engineer Konrad Zuse had seen the possibilities of computational intelligence in the late 1930s. But the Nazi government disregarded his lovingly hand-built constructions—a series of working electro-mechanical computers set up in his indulgent parents' Berlin living room. The apparatus was moved to Bavaria during the war and eventually carted off as war booty to Switzerland. After World War II, Germany was forbidden to dabble in electronics for at least a decade.

For a long time, the Soviets were bound both fiscally and ideologically. Ed Fredkin, then at MIT, once explained to me how computer programming was taught in the USSR. "It was like their swimming mandate," he said. "Everyone must know how to swim.

[9] Lively questions have grown up around the official finding of suicide. Turing had completed his humiliating "chemical castration" sentence. Although he'd lost his security clearance, he was engaged in important, non-secret research, and to his friends, seemed happy. He was known to be careless with the cyanide he was using in experiments. Thus his mother believed his death was an accident. Others have suggested he might have been murdered: he sat on some of the biggest secrets of World War II and, because of his homosexuality, was vulnerable to blackmail and other pressures. Lest he succumb to blackmail, removing him might be convenient. This seems farfetched because his homosexuality was no longer a secret.

Unfortunately, a desperate shortage of swimming pools made this impossible. So people were taught 'dry swimming.'" I see Fredkin leaning against a dark granite wall on West 116th Street in Manhattan, miming how to swim on dry land: he stands on one leg, kicking out the other, arms waving, a lampoon of the breaststroke. "Same with Soviet programming. Not enough computers for people really to learn. *Dry programming*." These circumstances provided precious little room for innovation, never mind the development of artificial intelligence. (Or, as a Soviet scientist once put it to Ed Feigenbaum, then at Stanford: "Who allows you to do this?")

These days AI is a thoroughly international endeavor and has been for decades. The Chinese, for example, intend to be world leaders, and the Japanese already are. Not incidentally, some of AI's most prominent scientists are women, making hash of an early accusation that the men creating AI were victims of womb envy.

To return to the possibly skewed AI foundation myth: the field's four founding fathers, John McCarthy, Marvin Minsky, Allen Newell, and Herbert Simon, stand as the four apostles (or horsemen of the apocalypse, depending on your point of view) of a reality we can all now see, a reality we all now inhabit. They saw this reality from the beginning. Yes, they were all Americans, but they would've been geniuses anywhere. The United States was wealthy enough, and its government leaders sufficiently visionary then, to allow their genius to flourish.

Thus one kind of AI was born through the brains and hands of a small brotherhood of scientists, all of them acquainted with each other, custom-crafting every program, laboring to make it all work on the primitive machines of their time. This story is partly about those people, most of them only spirits in my memory, who conceived that grand dream, that inevitability. They labored in what was then scientific isolation—often, scientific derision. AI was their way to understand human intelligence, possibly other intelligences, and they pursued it with glorious *joie de vivre*.

I won't have much to say here about the technical aspects of AI, which I wrote about at the field's dawn, in *Machines Who Think,* and which are ably described in several later excellent histories, textbooks, and survey articles. But from time to time, I'll look in on a line of research today, partly because one application or another intrigues me, and partly to bring the story up to date.

For each example I cite, please bear in mind that many similar research efforts are underway *around the world*. A full survey of AI today would need a study of encyclopedic proportions. To repeat Larry Smarr: this is no longer just a few programmers cobbling together Lisp programs. The whole world, every one of us, is at work on AI. We're all contributing, all doing our part, each time we go online; use our smartphones, credit cards, or social media; pass through automated toll booths; stream a movie; watch TV; you name it. We're all—if you worry—complicit.

Inevitably too, this story is about the people who felt threatened and were angry with me for being not only beguiled, but also sanguine. A pattern of what I call Dionysian eruption (after Nietzsche's distinction between the Apollonian and the Dionysian) characterizes the generally Apollonian history of AI. These Dionysian outbursts have often been ferociously passionate against AI, but sometimes, with equal passion, for it.

And this is about me, and my own journey through it all, as fascinated spectator, as accidental emissary between the Two Cultures of the humanities and the sciences. I'll say how it looked along the way and what I've learned over the years about thinking machines. I'll say why, as a humanist, I was drawn to AI, and where my intuition led. I'll say a little bit about myself so you can assess your narrator.

One thing I've learned is that humans can be serenely triumphal about extending our natural faculties of vision (eyeglasses, microscopes, telescopes); or locomotion (horseback, automobiles, everyday jet travel, space probes); or communication (writing, publishing, telephones, Skype), without ever being accused of tempting fate for our ambitions.

But extend our natural faculties of thinking? Illicit, sinister, blasphemous, hubristic. Anyway, impossible. (It's difficult to entertain that last notion any longer, but it obsessed many otherwise intelligent people for decades.) Reasons both obvious and subtle exist for all this, as you'll see.

Living in the exponential of AI, my great good fortune has been to watch most people evolve from jokey scorn to loud frustration that their computers and phones aren't smarter. I'm with that. Though I was its historian, kept its baby book, after some years I turned away from AI to other interests. For a while, I didn't pay much attention. When I turned back, intrigued by new programs, the moment was gravid.

AI's many subfields, such as machine learning, pattern recognition, vision, robotics, or natural language processing, once hived off like Protestant sects (and with some of the same moral indignation) but might begin to pull together to complement, interpenetrate, and amplify each other's purposes.

In these new ecumenical creations, human-level intelligence, or something even better, is thinkable.

3.

But let me be clear. An encounter with the Other has always brought with it very great uneasiness, especially when it concerns intelligence outside the human cranium. Western literature is full of this disquiet, from the Ten Commandments ("You shall not make for yourself a graven image, or any likeness of anything that is in heaven above, or that is on the earth beneath, or that is in the water under the earth…") to *Frankenstein* to *Neuromancer* to the daily news. With this disquiet I sympathize. Every grownup knows that technology giveth and technology taketh away. With AI, we aren't even far enough along on this path to be able to weigh the balance. Although I have misgivings, I take the long view and like to imagine what might be a better world if humans provide themselves with intelligent help.

All right, even intelligent computer overlords, in the words of Ken Jennings, the champion human *Jeopardy!* player who lost decisively but honorably to Watson: "I for one, welcome our new computer overlords." Deadpan, he alluded to an episode of *The Simpsons*, which probably borrowed it from Arthur C. Clarke's *Childhood's End.* (Watson would've got all that; I didn't.) We might all welcome intelligent overlords who—that—might save us from so much human folly.

At the very least, they'll bring us a fresh point of view.

But the strangest part of this sixty-year story is that for decades, I couldn't make otherwise intelligent and well-educated people believe that this could be important.

The Two Cultures

1.

My personal quest begins with a lecture that would shake and consternate the entire Anglo-American literary world in the mid-20th century, as surely as Martin Luther's Ninety-Five Theses shook the very foundations of the 16th century church.

In the autumn of 1960, the English department at the University of California at Berkeley had two visitors for the term. One of them was C.P. Snow, famous then as the author of the Strangers and Brothers series of novels about the British scientific world and the way mathematicians and scientists helped win World War II for Britain and the United States. The other visitor was his wife, Pamela Hansford Johnson. She too was a novelist, as well known as her husband, at least in Britain.

Snow and Johnson held an open house for students to come to tea each Thursday afternoon, and though I longed to go, I was much too shy. I was especially eager—but still couldn't bring myself to do it—after Snow delivered a talk about what he called "the Two Cultures," the humanities and the sciences, a kind of road show version of his Rede Lecture delivered a year or so earlier at

Cambridge University. This lecture was growing more famous and controversial by the month.[1]

The Two Cultures would reverberate so deeply—sometimes comically, sometimes painfully—through my life that it's worth saying a bit about Snow's theme. It outraged the First Culture, the humanities, into near apoplexy and foreshadowed what would become a needlessly Manichean struggle between human and artificial intelligence.

As he stood at the lectern in that autumn of 1960, Snow was a corpulent man in a baggy gray suit. His black-rimmed spectacles, his mouth, indeed his entire face seemed slightly too small for his

[1] Snow sketched the Two Cultures idea in a 1956 essay for The New Statesman and presented versions of it elsewhere before his 1959 Rede Lecture. He'd expand the ideas into long essays that appeared in two issues of Encounter in 1959. The fur flew—it got personal and nasty; British class prejudices were shamelessly exhibited to the embarrassment of American disputants, who thought one could disagree with Snow without resorting to such things. A small industry of pro and contra arguments labored furiously for years. It seems impossible to believe in the science-saturated world we inhabit now, but Snow's challenge was grievously heretical to the orthodoxy and intellectual hegemony of the humanities that had prevailed for almost three centuries in Britain and elsewhere (for the book was translated into many languages and sold well).Four years after the Rede Lecture, Snow published a reconsideration, whose major regret was that he hadn't used molecular biology instead of the second law of thermodynamics as a question to test scientific literacy. Stefan Collini's Introduction to Snow's The Two Cultures (1998, Cambridge University Press) offers a useful context for both Snow's arguments and those of his critics. Snow's grievance had a long personal genesis (fixed in the 1930s, Collini argues) and was based on both fact and passion, but it struck a public nerve, already raw from the work of the Angry Young Men, whose working-class, kitchen-sink realism (and anti-Modernism) was just then sweeping the West End and Broadway stages and soon Hollywood movies. The launch of Sputnik in early October 1957 had also disquieted the public. In 2008, the Times Literary Supplement named The Two Cultures and the Scientific Revolution as among the 100 books that had most influenced public discourse since World War II. As I walked across the Columbia campus in February 2015, I saw announcements of at least two lectures that had "Two Cultures" in their titles. Ed Feigenbaum recently told me that, as a post-doctoral student in 1959 England, he'd bought a first edition of The Two Cultures and still has it in a fireproof safe.

oversized bald head. His heft and his Oxbridge accent—adopted, I learned later—seemed deeply authoritative.

Two Cultures exist side by side, Snow began, the humanities and the sciences. Trained as a chemist and physicist and now writing novels, he was fortunate to travel in both worlds. But few others did. In truth, he went on, both our intellectual and our practical lives were dividing into polar opposites, the humanists and the scientists. Each group had lost the ability, much less the wish, to speak across the chasm. The humanists had taken the label "intellectuals" to describe themselves, but from that label they excluded scientists. In the view of humanists, scientists were shallow and brash (heroic age of science, *pah!*). Humanists were the sole custodians of culture, and culture meant only what they said it meant.

The scientists in turn believed that the humanists were "totally lacking in foresight, peculiarly unconcerned with their fellow men, in a deep sense anti-intellectual, anxious to restrict both art and thought to the existential moment," said Snow. The typical scientist thought that humanists were intellectually impoverished and vain about their ignorance as if traditional literary culture were the only culture that mattered. They behaved as if the natural order didn't exist, and its exploration was of no value, either for itself or for its consequences.

It was a caricature, he conceded, but not a very extreme one.

In 155 Dwinelle Hall, one of the biggest lecture halls on the Berkeley campus, I sat electrified by Snow's lecture. I'd enjoyed my required science courses. I'd thought seriously of declaring an anthropology or paleontology major instead of being an English major. I loved kneeling in the dirt, digging things out, and the possibility of finding human precursors seemed the ultimate thrill. What stopped me was not my comically romantic view of this enterprise, but one of the people I confided in. “How would you go on field trips?” she scoffed. “What would your husband and children do while you were away?” I had no answer. For the sake of my hypothetical husband and children, I kept on with English literature. Fine, I loved that, too.

Snow proposed a challenge: can you describe the second law of thermodynamics? A day later, John Paterson, leading my Virginia Woolf seminar, shook his head in—to his credit—bewildered amusement. “I don’t even know what the first law of thermodynamics is.”[2]

Snow’s ideas startled me, and at the same time put into words what, in some inchoate way, I already knew. Science students seldom showed up in my discussion-size classes, so I didn’t know any. I’d never dated an engineering student. What would we talk about? A course offered by the Berkeley philosophy department on British empiricism must address, I thought, the British Empire. I was basically ignorant about science, and sealed inside one of those two cultures, I’d never be privileged to break out. I regretted it, but it wasn’t a tragedy.

No. It was.

Fortunately for me, within weeks of hearing Snow’s talk, I was introduced to artificial intelligence, which began to shatter that seal. I didn’t make the connection between the two events at the time: that only appeared in retrospect, and only much later when I’d given up a

[2] The second law of thermodynamics says that in an isolated, or closed, physical system, disorder, or entropy, usually increases. It can never reorder itself. If Snow had mentioned that entropy is information concealed from us, or that information reduces uncertainty, as particle physicists believe, it might’ve added to the fun. That particular audience knew something about information.

yearning for human precursors and got involved with what might be human successors. But that connection would shape my life. It came about this way.

2.

I was working my way through college, a typist in the basement of South Hall, one of the last Victorian buildings on the Berkeley campus: tall, with wrought iron frills bedecking its steep mansard roof, aggressive ivy threatening its lofty windows. It housed the School of Business (then not nearly the glamorous and well-heeled academic discipline it would become).

From Summer 1959 to Winter 1961, I'd arrive every weekday afternoon from my English major courses—the Rise of the Novel, the Age of Milton, Modern Poetry, Literature of 20th Century France—and type course outlines for classes in administrative behavior, type reading lists for marketing, type midterm exams for decision-making under uncertainty, type scholarly papers in macroeconomics. One especially long summer, I typed an entire accounting textbook.

In those course outlines and reading lists, I briefly encountered the term *artificial intelligence.* But most striking, the name of Herbert A. Simon seemed to be everywhere. A course on municipal administration? The textbook was coauthored by Herbert A. Simon. Decision-making in organizations? Textbook coauthored by Herbert A. Simon. Theory of the firm? Herbert A. Simon. Introduction to artificial intelligence? Herbert A. Simon. It went on—and embarrassingly on. Fresh over from Dwinelle Hall and the riches of English and other European literatures, it was as if Shakespeare appeared in every course. (Or actually, Dryden. One of the minor greats you've heard of, but never read.) In my snooty English major way, I thought the field of business with all its little offshoots was surely the thinnest academic field ever, if one Herbert A. Simon had a hand in almost every textbook. I didn't know the word *polymath.*

Two of Simon's former PhD students had arrived in Berkeley from what was then Carnegie Tech (later Carnegie Mellon University) in Pittsburgh to be new assistant professors in the business school. Julian Feldman arrived in the fall of 1959, and Edward Feigenbaum, having spent a year on a Fulbright in England, arrived in that critical fall of 1960. Feldman always seemed rushed and overwhelmed with the press of life. (I was put to work typing transcripts of his psychology experiments with nonsense syllables, in this case, JIK and DAX, which gave me names for my two goldfish.) He was a big man and seemed to fill the room, not so much with his size but with his ever-harried presence.

Feigenbaum, on the contrary, was placid and soothing, no stereotyped hot-tempered redhead. He was always ready to stop and share a joke and, between puffs on his omnipresent pipe, shake his head slowly in sweet wonder at life. With his round face and benign grin, he seemed a bespectacled, sunnier, and infinitely more self-confident Charlie Brown.

Feigenbaum and Feldman were the emissaries of this new field called artificial intelligence, or as their part of it was called, computer simulation of cognitive processes, a fancy way of describing the effort to make a computer mimic certain aspects of human thinking. They were not only trying desperately to teach puzzled business school students, but also their colleagues all over the university.[3]

[3] Why this topic washed up in a business school at all is a bit of historical contingency.

The lack of textbooks compounded the difficulties of teaching AI to business students at Berkeley (and any others who found their way across the campus from psychology, operations research, or engineering). AI had been born as a distinct field only a few years earlier and only named by John McCarthy in the summer of 1956, when he and Marvin Minsky organized a summer workshop at Dartmouth under the guidance of Claude Shannon, the founder of information theory. None of these names meant anything to me then; some of them still don't mean much to the general public. But they and their colleagues would change the world.

3.

Feigenbaum and Feldman decided to put together a textbook of readings, so the most important AI work could be found between two covers instead of, in those pre-Internet days, hidden in the odd proceeding and journal. I was graduating in January 1961 and meant to go to law school the following September. But about the same time I heard C. P. Snow's Two Cultures lecture, Feigenbaum said to me, "Would you like to work on our book during that semester you're off?" I enthusiastically agreed. Only then did I ask what the book was about.

"Artificial intelligence," he said.

Herbert A. Simon and his former student, Allen Newell, had been teaching at Carnegie's Graduate School of Industrial Administration and were adventurous enough to go wherever their research took them. One major place it took them was the invention of thinking machines, computers that simulated certain aspects of human cognition. Anyone who wanted to study the new field with them had to enter Carnegie's Graduate School of Industrial Administration, and from there received his PhD. Feldman had originally trained in psychology and Feigenbaum in electrical engineering, so it was an odd but apparently happy path. Because they both had PhDs from the Graduate School of Industrial Administration, it must have seemed obvious to their colleagues at Berkeley that they'd fit in to the business school. They didn't quite, but no one yet understood that.

I needed a better definition than what I'd been able to glean from typing course outlines.

"Artificial intelligence," Feigenbaum said patiently, beginning the first of a thousand one explanations that, over a lifetime, would cover a grand list of topics, "is computers behaving in ways that if humans did that, we'd say, '*Ah!* That's intelligent behavior.'"

At the time, it was a fair answer. But in 1960, I must have taken a breath. I'd often been to the new computer center in another basement, a few buildings away from South Hall. With dire warnings from fretful faculty that I mustn't drop or mix up the stacks of punch cards, I'd deliver them across the counter to one of the clerks, who stood guard before the massive tape drives and processing units of the contemporary computer. Twenty-four hours later, I'd fetch the cards and whatever printout the computer had delivered as a result of running the program once. State of the art, 1960.

To associate all this in any way with intelligence was, well, ludicrous.

Intelligence was what the English, French, classics, anthropology, and architecture departments—all places I was taking courses—cherished as the *sine qua non* of human nature. Intelligence was central to literature, art, music, and beautiful structures. It had nothing to do with computers—all flashing lights and tape drives, on-off switches, dunderheaded insistence on flawless step-by-step instructions. Otherwise, like Victorian ladies, they'd swoon on their fainting couches, awaiting the smelling salts of debugging.

4.

Ed Feigenbaum's offer intrigued me. I needed a job during that break between graduation and law school. But something else was happeneing, too. As a literature major, I was thoroughly ground down by the pessimism of the post-World War II product, whether British, American, or French. Oppressed by it. Repelled.

I wasn't naïve. I'd gone through World War II—as a child, yes, but I'd gone through it, born in the middle of an air raid during the English Blitz, nothing between my laboring mother and the deadly night but a flapping tarpaulin. From my first breath, I was given to know that people in the skies wanted to kill me. I'd seen hard times afterwards, both in England and the United States, where my family immigrated. By the time I got to college, the Holocaust was the hottest topic of debate over long espressos at Il Piccolo on Telegraph Avenue. New friends had just fled from the Hungarian revolution. At the Cinema Guild and Studio, a storefront movie house on Telegraph Avenue run by Pauline Kael, one day to be the celebrated film critic of *The New Yorker* (Kael herself sometimes sold tickets and then disappeared to run the projector), I saw all the new European films. I could do world-weariness with the best of them.

But I was twenty. I was in love. Spring in Berkeley, summer in Berkeley, even autumn in Berkeley, were glorious with a sense of the imminent. Any season, the views across the campus to San Francisco Bay and the Golden Gate Bridge stopped you short with their panoramic grandeur. I'd begun to know that happiest of human moments, the simultaneous arrival of intellectual and sexual awakening. It was hard to be world-weary. No, it was impossible.

What I found among the artificial intelligentsia more than half a century ago was the same kind of excitement, optimism, and joyful hope in the future that I secretly harbored. But these men were open about it. They felt themselves on the cusp of something momentous. Grand. Epoch-making. They were right.

Thinking, Then and Now

1.

Ed Feigenbaum and Julian Feldman's book was *Computers and Thought.* The papers that went into this antediluvian volume of readings came from a variety of journals, such as *Proceedings of the Western Joint Computer Conference*, *Lernende Automaten*, the *Symposium on Bionics*, and *Proceedings of the Institute of Radio Engineers.* One came from the *IBM Journal of Research and Development*, and from *Mind* came Alan Turing's brilliant and amusing essay, "Computing Machinery and Intelligence," proposing the imitation game, which came to be called the Turing test.

Seeking these led me far away from the main library, where I'd done my own studying, to small science and engineering libraries scattered around the northern part of the campus. These early efforts in AI had appeared in so many odd niches because sometimes a scientist simply owed a paper to a journal or conference, and strategic publishing hadn't entered his mind. But really, the scattering signified how many researchers in different disciplines—engineering, psychology, business—were stirring with the possibility that these new computers might, in some way, be said to think. The zeitgeist was pregnant with the idea.

In the introduction to *Computers and Thought,* Feigenbaum and Feldman laid out their criteria for inclusion: the most important was that they wanted to focus on results, not speculation. *Results, not speculation.* Over the years, that issue has come up repeatedly because scientists from various fields used to believe (to use the words of one Nobel physicist I knew) they could "just come in and clean up AI." Although the field has taken various paths in its development and incorporated research from many fields, results are still the main requirement.

Feigenbaum and Feldman drew distinctions between what was then called *neural cybernetics,* where learning by computer would start from scratch (later known as *neural nets,* roughly, brain-like structures), and cognitive models. They favored cognitive models for two reasons. First, intelligent performance by a machine is difficult enough to achieve, they argued, without starting from scratch, as neural cybernetics required, from the cell up, so to speak. Therefore cognitive-model scientists built into their systems as much complexity of information processing as they understood and could program into a computer. Second, the cognitive model approach had yielded results whereas results from neural cybernetics were "barely discernible." This was to change, but not yet. Building into an intelligent system some intelligence already violated the philosophical notion of a mind as a tabula rasa, but it accorded with reality—humans are born knowing a lot and get trained as we develop.

What did I retrieve from all these arcane technical journals? Reports described programs that played chess and checkers—not brilliantly, but recognizably competing. Two programs proved mathematical theorems, and another program could solve symbolic integration problems in freshman calculus. Certain programs could answer questions (stringently limited in both topic and syntax) and others recognized simple patterns. Several programs had been written to imitate human cognition, at least as it was then understood, at the end of the 1950s. This too defined a division that would continue for a while in AI: on the one side, imitations—simulations, properly—of how humans think, and on the other, results achieved by whatever worked—mathematical models, statistical models, algorithms.[1] In those early days, the field of AI was ecumenical.

Human intelligence seemed as obvious and substantial as the Great Wall of China, but when humans reached out for it, it dissolved in a miasma of conjectures and swamp gas. Computer programs might offer a way of modeling and *understanding* it by actually behaving intelligently for anyone to see.

I fetched, cut, pasted, and typed. Because my two bosses didn't care when I worked so long as I got the work done, I mostly worked evenings, when my new husband, Tom Tellefsen, was absorbed as an architecture student. Typical of computing, nighttime was when the lab and office came alive. As I worked, I was surrounded by young graduate students in the field who were not only high-spirited good

[1] Decades later, I've heard people say that everything in the field was laid out in principle in Computers and Thought, and the subsequent work is mere technological commentary. No. AI has evolved mightily in the sixty years since the book was published thanks to new ideas, new techniques, and dramatically better technology. The book doesn't include a word about robotics for example. Certainly many fundamental ideas appeared in Computers and Thought, but they've developed in unforeseen ways.

company but also taught me by osmosis things I needed to know about the scientific method.

Gradually the book was assembled, with Feigenbaum and Feldman adding brief paragraphs that gave the context of each paper. It seemed a natural for the Prentice-Hall series in computing, but the consulting editor of the series, the ubiquitous Herb Simon, told Prentice-Hall that the book wouldn't sell, and they should reject it. (For that, Simon would laugh at himself for many years to come. More than a half-century later, the book is available for nothing on the Web, but The MIT Press will sell you a printed and bound copy if you wish.)

McGraw-Hill signed it up gladly. The firm had set up a branch office in Marin County, across San Francisco Bay from Berkeley, and I have vivid memories of going with Feigenbaum and Feldman to visit their editor. I'd read books since I was four. I had, in some very vague way, an idea that I might be involved in writing books someday. A visit to an actual publisher? Catnip.

By the time *Computers and Thought* was published in 1963, I was gone from the campus. I hadn't entered law school after all but was working instead in my family's business. It wasn't a good fit for me, so I was thrilled when, in 1965, Ed Feigenbaum called and asked me to join him at Stanford as his assistant. He'd finally thrown over his dispiriting missionary work among members of the Berkeley faculty and decamped to Stanford, which had an actual computer science department, one of the first.

2.

"It can't *think* because it's not *human*!" a dear friend shouted at me recently. He meant a computer. In the words of Harvard's Leslie

Valiant, my friend is confusing what the computer *is* with what it *does*. (Valiant, 2014). But my friend has plenty of company in his conviction.

Yet in the last half century, something has changed. These days, what we consider to be intelligence, thinking, or cognition has stretched to encompass much more behavior and extended to many more entities than anyone in the 1960s could have anticipated. Animal behaviorists study intelligence in primates, cetaceans, elephants, dogs, cats, raccoons, parrots, rodents, bumblebees, and even slime molds, and nobody now is surprised. Entire books appear on comparative intelligence across species, trying to tease out what's uniquely human. It isn't obvious. Our fellow creatures are pretty smart. David Krakauer, a theoretical biologist and president of the Santa Fe Institute, an independent think tank, argues that in biological systems cognition is ubiquitous, from the cell on up, from the brain on down. But so far as we can tell, no other animal seems to possess the faculties of our uniquely wired frontal lobes, the seat of planning, self-restraint, elaborated language, and symbolic cognition. As we'll see, that competence has astounding effects.

In 2013 Dennis Tenen, a young assistant professor of English at Columbia University, presented a humanities seminar at Harvard. He suggested that maybe intelligence should be considered to reside in systems, not in individual skulls, and wasn't hooted out of the room. However novel this idea is to humanists, it's been central to AI for decades. Even humanists need look back no further than "Among School Children" by William Butler Yeats:

> O body swayed to music, O brightening glance,
>
> How can we know the dancer from the dance?

Feigenbaum's explanation to me, back in 1960 ("*Ah!* That's intelligent behavior") was only an operational definition of intelligence. He hadn't said that a computer needed to imitate *human* thinking processes. We simply had to recognize its behavior, or output, as what we'd call intelligence if humans did it. Of course this itself is problematic, specific to time and culture. In the 19th century, clerks and bookkeepers were paid professionals; their jobs required intelligent behavior. Without much fuss, machines have long since replaced them. Later, as programs grew more complex, Feigenbaum would add that any computer doing a task that required intelligence needed to be able to explain its line of reasoning to human satisfaction.[2] This idea, early articulated by Feigenbaum, has re-emerged compellingly as flash algorithms make decisions, sometimes life and death, which no one can verify.

3.

More than half a century after Ed Feigenbaum's brief 1960 definition to me, Russell and Norvig (2010) grouped AIs into four general categories:

acting humanly;

thinking humanly;

thinking rationally;

and acting rationally.

[2] A 2013 Google system seems to have baffled everyone. During an exposure to millions of random YouTube videos, it unerringly identified images of paper shredders. Unlike cats, paper shredders are an object most humans can't identify and whose features Google's engineers were unable to classify and code with any precision. The system isn't saying how it learned. In response to Feigenbaum's declaration, some decades later, others now support this need for explanation, so some deep-learning programs are slowly being equipped with explanatory lines of reasoning.

Acting humanly means the artifact can hear and speak a natural language, store what it knows or hears, use that knowledge to answer questions and draw new conclusions, adapt to new circumstances, and detect and extrapolate patterns. It might also be able to see and manipulate objects. It would know the rules of social interaction—if embodied, the right distance to stand from humans, how to conduct a conversation, when to smile, when to look solemn. No single program, no artifact extant, can do all these things now.

Thinking humanly means to model specifically human ways of thinking, drawing on (and contributing to) the latest in cognitive psychology and neuroscience. This is a way to understand human cognition, whether planning tasks, interpreting a scene, or lending a hand. With a sufficiently precise theory of some aspect of mind, a computer program can express that theory. Imitating input and output isn't enough: the program must trace the same steps as human thought and be part of the construction and testing of theories of the human mind.

Thinking rationally is to obey the laws of thought, reason, or logic, sometimes expressed in the formal terms of logical notation. An Aristotelian ideal of thinking, thinking rationally often runs into trouble in the messy real world.

Acting rationally refers to an agent—something or someone—operating autonomously in its environment, persisting over a long time period, adapting to change, and creating and pursuing goals. A rational agent acts to achieve the best outcome (or in uncertainty, the best *expected* outcome: a rational agent isn't omniscient, perfect).

Rational agents offer two advantages. First, they can be more general than the "laws of thought" approach allows. Second, rational agents are, in the words of Russell and Norvig, "more amenable to scientific development than are approaches based on human behavior or human thought." That's a tactful way of saying rational agents can improve quickly. Humans take awhile, sometimes never. These days, most AIs can be characterized as rational agents. As defined here, perhaps acting rationally and acting humanly are slowly converging in computers. But it hasn't happened yet. Nor in humans, come to that.

A second look at those categories, acting and thinking humanly, acting and thinking rationally, and I'm struck by how intimate this machine is. With AI, we're creating our own doppelgängers, or a new, improved version of humans or maybe our successors. "Just tell me," the machine seems to say to us softly, flattering as a lover, "Just tell me how you do it." Irresistibly, we move closer. "So that I can do it too." We back away. "Just tell me." We move closer again, of course. We confide, impeded only by what we don't know, our own tricks we have no conscious access to, which turn out to be deeper and more complex than we expected. I'm also struck by how often *humanly* appears in these four definitions. That reflects a long-held belief that intelligence is solely a human property, although as I noted earlier, in recent decades, scientists have widened the definition of intelligence to include other species.

Perhaps physicist Max Tegmark's distinctions in *Life 3.0* (2017) are useful here. He cateogorizes three stages of life: biological evolution, cultural evolution, and technological evolution. "Life 1.0 is unable to redesign either its hardware or its software during its lifetime: both are determined by its DNA, and change only through evolution over many generations. In contrast, Life 2.0 can redesign much of its software: humans can learn complex new skills—for example, languages, sports and professions—and can fundamentally update their world-view and goals. Life 3.0, which doesn't yet exist on Earth, can dramatically redesign not only its software, but its hardware as well, rather than having to wait for it to gradually evolve over generations."

What I didn't understand until long after I'd followed my twitching nose down the AI path was that the computer would be the instrument that finally brought the Two Cultures together. AI and its principles would be central to that rapprochement. Given the thrashes I'd get into over the years trying to explain AI, and computing generally, to my colleagues in the First Culture, nobody else understood this either.

Such understanding dawned gradually. By the early 21st century, a field called the digital humanities had blossomed, although that's only the most obvious sign of detente. I also mean something deeper, both intellectually and emotionally—the beginning of that enormous scientific enterprise called computational rationality that I described earlier, to encompass, explain, and account for intelligence in all its guises.

Learning a New Way of Thinking at Stanford

1.

In 1965, the Stanford University campus was surely one of the loveliest places in the world. My father once read—and never forgot—that on all Planet Earth, the ideal daily temperature for humans is in Redwood City, California, a few miles from Stanford. Days are sunny and warm; cool evenings come as the marine fog creeps over the Santa Cruz Mountains in the late afternoon from the west. For variety, it rains a little, but mostly the sun-drenched days follow each other benevolently; Stanford students and faculty bike (and nowadays skate and skateboard) everywhere.

From the old Southern Pacific Railway station in Palo Alto (I was living in San Francisco then), I biked daily beneath the noble Canary Island date palms of Palm Drive, an allée that led to the golden sandstone mission-style main campus. Then I'd swerve north to Polya Hall.

George Pólya had been the great mathematical star of the Stanford campus since 1940, when he left ETH, the Swiss Federal Technical Institute. He'd made significant contributions to many mathematical fields and was still an active emeritus professor. He was particularly fascinated by the study of *heuristics*—a word he popularized. Just how did people go about solving problems? What rules of thumb did they use? This would be dear to the hearts of people in AI, who believed that heuristics compensated for computing deficits in the human brain.

Polya Hall was temporary. The firm of Joseph Eichler, best known for his residential subdivisions that were modernist architecture for the masses, had sited a small group of buildings around an ancient, gnarled live oak, a tree typical of the California coastal range. Except for Polya Hall, which had two stories, these were single-story white stucco with prominent beams and floor-to-ceiling windows that welcomed the California light. Computers might not care about sun and light, but the humans who used and tended them certainly did.

My office in Polya Hall had a door connecting to Ed Feigenbaum's and looked out on a pretty side garden. Until I got a study overlooking Riverside Park and the Hudson River, it was the most appealing office I'd ever work in. It was a joy to arrive in the morning, the fog newly burned off to reveal indigo skies, the air fresh, the flowers blooming year-round, and get to work.

Along with his professorship in the new department of computer science, Feigenbaum had also been appointed head of the computation center, a computing utility for the entire campus. As his assistant, one of my first projects was to interview faculty users all over the campus and ask what they used the computation facilities for. After I compiled and wrote up that information, the computation center decided it was time to replace its aging (maybe two years old) Burroughs B 5500 Information Processing System with a new system, the IBM System/360. This was next-generation computing: it was up-and-down compatible, meaning the machines with the smallest memories and processors would still be compatible with machines at the high end. It could also combine two kinds of computers that had been separate lines at IBM, those for business and those for science and engineering.

Other manufacturers of high-end mainframes mounted a lively competition and courtship. I took minutes at meetings where claims were staked, promises made. But Ed and his colleagues decided that IBM offered the best solution to their problems, so they chose the S/360. Alas, like a princess unready to meet the prince signed to the marriage contract, the new system was coy. Software problems, hardware problems—the later term was *vaporware.* I'd struggle to keep a straight face as the IBM reps swept in like Mafioso dons (although far better dressed—conspicuously well-cut suits and rich ties on the decidedly informal Stanford campus). They managed to swagger and apologize simultaneously, then finish with, "But that's history. What can we do for you now?" An unforgettable response. Later I'd find an occasion or two to use it myself.

Ed would assess them in silence as he puffed on his pipe. Then he'd say evenly, "When do you think we'll see the system?"

"We're working round the clock," the IBM reps would assure him eagerly. Ed would remain quiet, puff, blink thoughtfully behind his horn-rimmed glasses. After a silence, he'd respond, "Let's see something."

Exasperating as this whole episode must have been for him, I never saw Ed lose his equanimity. In some sixty years of friendship, I've never heard him raise his voice. This was another remarkable lesson for me. I'd grown up in a family of Anglo-Irishmen who understood the tactical value of bellowing, voices that rattled the windows in their frames and made female knees knock. Many years later, I whimsically wondered whether one of the great attractions of AI for me was its antipodean location from the rage and unreason that, as a child of the mid-20th century, had surrounded me from the beginning, both on a world scale and in the domestic theater of my family.

The world's rage and unreason needs no repeating, but in my family, rages, tantrums, sulks, threats, and high-decibel shouting (my father) led to rebellious but submissive tears (my mother and us kids) at least weekly. To a soul that longed for serenity and a taste for intellectual stimulation, AI, based on calm reason, was emotional balm. Yet into AI's cool Apollonian order, the Dionysian too would suddenly erupt: it's always with us.

This description of my family's domestic theater is incomplete. It was richer in comedy than melodrama, rich in music as well as words. My mother, a country girl from an English village so picturesque it belonged pictured on the top of a biscuit tin, was a gifted musician. My first real memory, as distinct from ones told to me, was of sitting in our front room and listening to her play Debussy's *Clair de Lune* on our piano. She'd been an independent career girl in the 1930s, a busy hairdresser by day, pianist in a dance band at night, and her glamorous sequined crepe gowns showed up in my dress-up box years later. She was hard-working, generous-minded, enthusiastic, fascinated by humans in all their variety. I see now she paid a heavy price for the 1950s backwardness of women's roles, especially in the suburbs of California where we fetched up. Later, her kindness and sense of humor (she was known in her seventies and eighties as "Sunshine" to her fellow musicians), along with her renewed devotion to music and dancing, saw her admirably through twenty-five years of merry widowhood. She was a witty storyteller about her fellow musicians and about astonishing family lore. My father dead, she and I became friends in a way we couldn't have earlier. In that late mother-daughter friendship, I began to see how I owed some of the best of myself to her.

My father was ambitious, intelligent, and eventually a very successful businessman in the New World; he was handsome, suave, utterly charming, with an abundant Irish storyteller's gift. Thus he'd transmute stories of his grievously poor boyhood, moments of excruciating humiliation (I now see) into tales so funny that we wept helplessly with laughter and begged to hear them again and again. The boy in the pith helmet (castoff from some relative in the colonial service) provoking the Liverpool schoolmaster to cry, "Willy McCorduck! Come out from under that hat!" Or the men's trousers so inexpertly cut down that the flies reached to his bony knees. Or how the nearly mythical aunts in London took him over and taught him manners, and he almost starved when he came back to the anarchy of his own family's Liverpool dinner table. There were dozens of these stories; we could almost mouth the words as he spoke them, we'd loved and heard them so often. We loved *him*. Thanks to the comedy, and vibrant intellectual curiosity that always drove him, we forgave the rages and tantrums. But we couldn't forget them.

2.

My Stanford office was a gathering place. Anyone who wanted to see Ed Feigenbaum, whether in his role as professor, or as head of the computation center, had to come through my office. Waiting, they chatted amiably with me, which is how John Reynolds taught me about Venn diagrams and Joshua Lederberg, the Nobel Laureate in genetics, gave me a quick course in mass spectroscopy.

In retrospect, it seems unbelievable that at the same time Ed was running Stanford's computation center and dealing with recalcitrant computer manufacturers, he and Lederberg were starting research with the audacious goal of studying how scientists formed hypotheses. This project would turn AI research upside down and overthrow convictions of more than two thousand years of Western philosophy.

I didn't really understand the impact of this research until ten years later, when I began writing my history of AI, *Machines Who Think*. As I'd discover, the program called Dendral, led by Feigenbaum and Lederberg, involved the radical rethinking of induction, human as well as machine. Its emphasis was securely on knowledge, not reasoning. I'd later come to think of Dendral as the Tristan chord of AI, a bold insight that changed everything.[1]

So, like the Tristan chord, knowledge was present in early AI—the rules of chess, for example—but Feigenbaum and his colleagues changed the emphasis radically: knowledge, not reasoning. This emphasis makes more sense if I consider Dendral in context when, later on, I write about the early days of AI.

[1] In Richard Wagner's opera, Tristan und Isolde, the Tristan chord is a four-note mix of major and minor thirds, which was shocking at the time, although some musicologists argue it wasn't original with Wagner. Its shock came from the emphasis Wagner gave it. That chord later found a welcome home in the American Songbook—Arlen, Ellington, Gershwin, Strayhorn—whose fusion of late Romantic tonality and African-American rhythms (not to mention a story in 32 concise bars) leapt out of Tin Pan Alley and captivated the world.

3.

My husband Tom and I were weary of San Francisco's perpetual fog and wind. We realized we could live in that ideal climate of the sunny, warm peninsula that had seized my father's attention when he read about it so many years earlier. We found a tumbledown little cabin to rent, a mile or two west of the village of Skylonda, the crossroads of La Honda Road and Skyline Boulevard, high above and west of, Stanford. The cabin clung precariously to a canyon wall on the western slopes of the mountains. (It literally slid into the canyon some years after we left.) We heard it had once been the local brothel, but our neighbor laughed. "Oh no," he said, "speakeasy. Isn't that big bar still in the cellar?" It was.

Now, instead of taking the old Southern Pacific train and then my bike, I could zip east up La Honda Road, over the Skyline Boulevard summit, and down to Stanford in my black Volkswagen Beetle, which lacked both gas gauge and seat belts.

The cabin was isolated, as different as possible from our former Potrero Hill flat above the Bayshore Freeway in San Francisco—where, for a stunning view of the San Francisco skyline, we inhaled ominous levels of traffic fumes. In the mountains, we were surrounded by lofty coast redwoods and tan oaks and stepped over fat yellow banana slugs every morning as we made our way up the packed dirt stairway to our cars. We knew our neighbors by sight but at night couldn't even see their lights.

Further down La Honda Road, Ken Kesey and his Merry Pranksters lived and partied in the village of La Honda itself. I'd smile to pass the Pranksters' psychedelically painted bus parked at the corner of Skyline and La Honda Road, in the garage for repairs almost perpetually. Every weekend, sometimes weeknights, the Hell's Angels would shatter the forest silence, their Harley hogs in formation down the mountain road on a visit to the Pranksters. Tom was often in the city, working and taking night courses. I began to feel vulnerable. Once, I thought I had an intruder and called the sheriff's office. The officer finally arrived and shook his head. No way anyone from the sheriff could get to us in less than half an hour. He strongly recommended I arm myself. Tom and I went to Sears and bought a hunting rifle, no ammunition, and this we stored under the bed. I only knew from the movies which way to point it.

It never occurred to me to go down the road and introduce myself to Kesey and his Pranksters. I was shy, serious, and, I thought, probably much too straitlaced for their wild ways. Dope didn't interest me; my eccentricities were still seedlings. When I met some Pranksters much later, we liked each other, but we'd all transformed, neither wild prankster nor prim Stanford employee.

Two decades on, I met Kesey. My literary agent then, John Brockman, ran occasional meetings called The Reality Club, and we'd gathered to honor Kesey, who was in town to promote a book. We met in somebody's immaculate white apartment, high above Central Park, with a southern view to the dazzling luminosity of nighttime midtown Manhattan. I introduced myself to Kesey and told him how our paths had almost crossed twenty years earlier. He looked around this magazine-shoot apartment and began to laugh a deeply amused, generous laugh, reached out, and squeezed my hand. "You've come a long way, baby!"

The La Honda cabin made me self-sufficient. It had a pitiful and expensive propane gas heating system, so its fireplaces were the main source of heat—up in the Santa Cruz mountains, we had cold nights even in midsummer, when the Pacific marine fog flowed into the canyons and up to the mountain summits. We had occasional snow in the winter. So I learned to split wood. I learned to crawl, hand-over-hand, across the face of the canyon wall to find and fix the break in the water system, probably a monthly event. We got a dog, a Saint Bernard, and I was a spectacle when I bundled behemoth Sebastian into the back seat of my Beetle and tooled around Palo Alto.

Living on the peninsula also meant that I could spend some evenings and weekends working on my writing, using the luxurious electric typewriter that I used all day in the office. But many people worked evenings and weekends in Polya Hall, and if luck was with me, somebody would drop by and interrupt—"You doing anything important? No"—and chat. My favorite visitor was Bill Miller, a specialist in interpreting the graphics of bubble chambers for basic physics research. He was later Stanford's provost and later still, CEO and president of the Stanford Research Institute. Miller was simultaneously focused, far-sighted, and practical.

His vision is apparent in what he went on to accomplish, not just as Stanford's provost and vice-president, but in the major role he played in continuing to nurture—on those acres of former cherry, apricot, and peach orchards—what Frederick Terman had begun, what came to be known as Silicon Valley. In addition to his academic roles, he was an early venture capitalist, an entrepreneur, and board member of many organizations. Above all, he was someone who saw and made others see the social, scientific, and economic promise in computing.

Miller was born in rural Indiana and went to university nearby at Purdue. His open and planar face made him appear a bit simple, which he decidedly wasn't. He had a farm boy's friendly skepticism combined with utter pragmatism, which colored everything he did. He thought much about how people behaved and had begun to coin pithy precepts, presented to me as the "The Sayings of Middle-Class William." Always negotiate, Middle Class William counseled. If you negotiate from strength, you have nothing to lose. If you negotiate from weakness, you still have nothing to lose. His tales of boar hunting when he'd been a young officer in the American Occupation in Germany found their way, decades later, into one of my novels, *The Edge of Chaos*. I don't know if it was his army or his farm experience that inspired him to improvise support for a young computer music composer, John Chowning. Miller told funders that Chowning was studying the ambient noise in computer rooms.

Miller was full of practical advice. When I needed a cat to keep the mice in the cabin manageable, he said, "Be sure and get a kitten whose mother is a mouser. Mousing doesn't come naturally to cats; they need to be taught." But if I asked him how he liked living in all this glorious Palo Alto sunshine, he'd be puzzled. He'd look out the window—yes, there it was—and shrug. "I might just as well be in Indiana," he'd say. "My work is what I care about."

No words could've surprised me more. I'd always been deeply sensitive to place. When I arrived in California as a seven-year-old, I emerged suddenly into its brilliant light out of a great, oppressive darkness that had literally silenced me. You don't think of graves at that early age, but I knew the California light had restored me to life.

4.

At Stanford, I was learning by osmosis again, the way I'd learned from the graduate students at Berkeley. I was mainly learning about

AI, deeply important at Stanford, which, along with Carnegie Mellon and MIT, was then one of the three great world centers of AI research. All three were undisputed world centers of computing research generally, and it's no coincidence that AI was centrally embedded in that wider, pioneering research.

Ed Feigenbaum had come to Stanford hoping that he and John McCarthy could collaborate. They remained personally friendly but realized their destiny was to pursue different paths in AI research. When I arrived at Stanford, McCarthy was in the process of moving his research team to a handsome, low-slung semicircle of a new industrial building in the Stanford hills, perhaps five miles from Polya Hall. A now defunct firm called General Telephone and Electric, seeing the new structure didn't fit their research plans after all, had given it to Stanford, and it became the Stanford Artificial Intelligence Laboratory, SAIL.[2]

Among the research projects that had moved from Polya Hall to SAIL was Kenneth Colby's Doctor program. Colby was an MD and psychiatrist who thought there must be some way to improve the therapeutic process—perhaps by automating it. Patients in state psychiatric hospitals might see a therapist maybe once a month if they were lucky. If instead they could interact with an artificial therapist anytime they wanted, then whatever its drawbacks, Colby argued, it was better than the current situation. In those prepsychotropic drug days, Colby wasn't alone in thinking so. Similar work was underway at Massachusetts General Hospital. Colby had collaborated for a while with Joseph Weizenbaum, an experienced programmer, who'd come from a major role in automating the Bank of America

[2] Read an autobiography of SAIL at http://infolab.stanford.edu/pub/voy/museum/pictures/AIlab/SailFarewell.html

and was interested in experimenting with Lisp. Weizenbaum would soon create a dialect of Lisp called Slip, for Symbolic Lisp, though it really had no symbolic aspirations as AI understood the term.

Doctor was the program that the visiting and eminent Soviet scientist, Andrei Yershov, asked to see. His encounter with Doctor was the moment that artificial intelligence suddenly became something deeper and richer for me than just an interesting, even amusing, abstraction.

But Doctor raised questions. Should the machine take on this therapeutic role, even if the alternative was no help at all? The question and those that flowed from it deserved to be taken seriously. Arguments for and against were fierce. Weizenbaum warned that the therapeutic transaction was one area where a machine must not intrude, but Colby said machine-based therapy was surely preferable to no help at all.

Thus Doctor was the beginning of a bitter academic feud between Weizenbaum and Colby, which I would later be drawn into when I published *Machines Who Think* and made for myself a determined enemy in Weizenbaum.

At the time Yershov was playing (or not) with the Doctor demonstration, Weizenbaum was already beginning to claim that Colby had ripped him off—Doctor, he charged, was just a version of Weizenbaum's own question-answering program, called Eliza (after Eliza Doolittle). Eliza was meant to simulate, or caricature, a Rogerian therapist, which simply turned any patient's statement into a question. *I'm feeling sad today. Why are you feeling sad today? I don't know exactly. You don't know exactly?* Feigenbaum, who'd taught Lisp to Weizenbaum, says that Eliza had no AI aspirations and was a no more than a programming experiment.

Colby objected strenuously to Weizenbaum's charges. Yes, they'd collaborated for a brief period, but putting real, if primitive, psychiatric skills into Doctor was Colby's original contribution and justified the new name. Furthermore, Colby was trying to make this a practical venture whereas Weizenbaum had made no improvements in his toy program.

Maybe because Weizenbaum seemed to get no traction with his claims of being ripped off, he turned to moralizing. Even if Colby could make it work, Doctor was a repulsive idea, Weizenbaum said. Humans, not machines, should be listening to the troubles of other humans. That, Colby argued, was exactly his point. Nobody was available to listen to people in mental anguish. Should they therefore be left in anguish?[3]

I agreed with Colby. Before this, I might not, and based only on first feelings, have sided with Weizenbaum.

But at Stanford, I was learning to think differently. One day, I tried to explain to Feigenbaum how I'd always groped my way fuzzily,

[3] By 2018, online therapy was thriving. One project, a joint effort between Stanford psychologists and computer scientists, and called Woebot, offered cheap but not free therapy to combat depression. It was a hybrid—one part interaction with a computer, and one part interaction with human therapists, this for people who couldn't afford the high cost of conventional therapy. Earlier projects included one at the Institute for Creative Technologies in Los Angeles, called Ellie, to assist former soldiers with PTSD. Ellie's elaborate protocols seem to have overcome the problem that many patients resist telling the truth to a human therapist but feel freer with a computer. (We saw this with Soviet computer scientist Andrei Yershov.) Some decades ago, the Kaiser Foundation discovered the same reaction to ordinary medical questions—people felt judged by human doctors in ways they didn't by computers and could thus be more candid.

instinctively into issues, relying on feelings. Now I began to think them through logically. Ed laughed. "Welcome to analytic thinking."

I'd entered university hoping to learn "the best which has been thought and said in the world," as I read in Matthew Arnold's *Culture and Anarchy* in my eager freshman year. But Arnold said more: the purpose of that knowledge was to turn a stream of fresh and free thought upon our stock notions and habits.

For me, meeting artificial intelligence did exactly that.

5.

Other things were happening in my life. Evenings, I'd been taking writing workshops at the University of California Extension in San Francisco, led by an ebullient New Yorker called Leonard Bishop, a successful novelist in the 1950s and 1960s. He took me aside and told me, in his delicious Lower East Side bawl, that I'd better start taking all this writing stuff seriously, awright?

I was restless. I was ambitious to be something beyond somebody's assistant. When the computation center decided it was generating enough documents to need a technical writer and editor, I applied for the job. Ed was upset, he told me later, but instructed the man doing the hiring that I should be considered just like anyone else. I didn't get that job—it went to a slack-jawed guy I called The Schlump (along with AI, I was learning Yiddish) who, for all I know, was better qualified. But I couldn't keep on as I had. My marriage was dissolving. I decided to get away from it all and go east.

One of the last days at Polya Hall, I stood gazing out at the computing quadrangle, at the great old live oak at its center and watched the graduate students come and go among the various buildings. There was Pat Suppes's early project on teaching machines, the computation center that housed the mainframes, and an adjacent building that was home to a mathematics education project. I knew nearly all of those graduate students. The word was barely in use, but nerds are what I suppose we were. Geeks.

We did know that we were all kin in a strange point of view, a fellowship, a clan, maybe a cult, which nearly nobody else in the world was aware of. In 1966, not all of us could possibly grasp the radical changes, the immense profundity, of the coming information revolution, where nearly everything, from biology to physics to literature to commerce to the arts could be described in information processing terms, where the world was going to go digital. That early in the revolution, we could only take soul-nourishing comfort in connecting with other members of our rare tribe. But we were utterly confident, unshakably sure, that what we were devoted to would somehow change everything. As I headed back to the outside world, I wondered if I'd ever feel so at home again.

If I was so happy living in the future, what drew me backward to the present, the past, that outside world? It was the gravitational pull in my mind of the First Culture—literature, the humanities—that I still imagined was somehow primary. This was not a culture that would ever particularly welcome me, nor appreciate what I was saying. But for a long time, I'd yearn for it just the same.

Revolution in the Rust Belt

1.

A job waited for me in New York City, and I started at Columbia as a part-time writing student. In the autumn of 1968, thanks to a generous fellowship, I could go full-time in the writing program at Columbia's School of the Arts. Age 27, a marriage behind me, I lived at International House, happier than I could remember. I had nothing to do but write.

The writing program then required its students to design a course of study that would complement the writing workshops (so we might be gainfully employed between royalty statements?). I designed a program to study human factors in computing, surely so exotic-sounding that in 1968 nobody at the School of the Arts had the remotest idea what I meant to do. So what? Alongside my writing workshops, I took courses all over the Columbia campus with a joyful heart.

I've made my share of snide remarks about "the workshopped novel" in its predictable emollience, but those two years at Columbia's School of the Arts were an unsurpassed gift to me. I was surrounded by writers whose work I knew—Adrienne Rich, Jean Stafford, Stanley Kunitz, Frank MacShane—and by young writers whose work I'd come to know. Visiting writers brought in the fairy dust of celebrity: Reynolds Price, Robert Penn Warren, Jorge Luis Borges.

I was blessed to be the continuing student of Hortense Calisher, who subsequently claimed she mostly left me alone. But years later, I found one of my manuscripts thoroughly marked up. Calisher and I began as student and teacher; we became dear and loving friends for forty years until the end of her life.[1] I was also the student of V. S. Pritchett, a visiting British writer, and thought I'd never known a kinder man, deeply astute about writing, but who understood in his bones how fragile a writer can be. He took my first novel, *Familiar Relations*, my nominal masters' thesis, and gave it to his own agent, who immediately found a publisher for it in London, Michael Joseph. Victor Pritchett and I corresponded for several years after I was gone from Columbia and he was back in London.

I fell in love again. With two men simultaneously. I was swept up by a witty Berlin judge, also living at International House, in the United States on a special fellowship to study the American primary system. I was enthralled by Joe Traub, a computer scientist at Bell Labs, whom I'd met at Stanford, but got to know in New York City. I was having the grandest of times.

[1] Hortense Calisher and I had many reasons to be friends, but among them was her understanding and admiration of C. P. Snow's work. She loved the idea of creating in a series of novels, Strangers and Brothers, a significant living world that most of us wouldn't otherwise be permitted to enter, the world of big science and its role in government, especially in winning World War II. She admired his larger ideas of what being intellectual encompassed. In her memoir, Herself, she describes spending a day with Snow and his wife, Pamela Hansford Johnson, hating to tear herself away. Decades later, she wrote a long novel called Mysteries of Motion, which had original and provocative things to say about the space program, including what she called "the odd fastidiousness of intellectuals who believe it has nothing to do with them."

Ed Feigenbaum came to visit me sometime during that year, amazed at my transformation. Well, yes. Skinnier, in mini-skirts, tights and high leather boots, doing what I loved, growing in self-confidence. “It’s like being rich,” I confided to him. “You can do anything you want and get away with it.”

In June 1969, at the end of the school year, the judge went back to Berlin. We were deeply serious about each other, but as a writer I couldn’t bring myself to leave my mother tongue. My growing feminism puzzled and worried his orderly German soul. He and I would continue to be friends for the rest of our lives, seeing each other (and our spouses) in New York City, which he loved, or his Berlin, which I came to love.

But Joe admired my independence, even when it upended everything he’d been taught about the relationships between women and men. I moved in with him in the Village, and we married that December, when he persuaded me that we could save in taxes if we married within the calendar year. It wasn’t really the money. I had an even more generous fellowship my second year at Columbia, thanks to Delacorte Press. Tax incentives were just an excuse.

I loved Joe for many reasons. I struggle to say which matters to me more: his steadfast support of my work, his deep abiding kindness to me (a much underrated characteristic in a marriage), his merry heart, or that he never bored me. For nearly fifty years, he came to the dinner table with a list, often written, of topics he wanted to amuse, stimulate, or provoke me with. Sometimes I’d get an email tease: “Two star item over dinner tonight.” I’d have to wait until he got home and dinner was on the table to hear it. He understood, nourished, and shared, all of my deepest passions.

I loved Joe's take on life, which I suppose is a scientist's take: question the givens, take nothing for granted, ask the awkward questions, look at it all differently. Many years ago, I gave him a Christmas present, a hefty river stone with a single word carved on it by the stonecutters of The Cathedral of St. John the Divine: WONDER. This word summed him up for me, both his eternal, childlike wonder and delight at the world, and the word that started most of our conversations: "I wonder…" We put the stone in a small garden outside his study in our Santa Fe house, a place where he loved to read, look at his garden, oversee his birds, and lift up his eyes in joy—and wonder—at the Sangre de Cristo Mountains.

Although he died suddenly in 2015 in that house he loved, surrounded by those mountains he loved, died in my arms, he's still alive to me in my heart, my mind, his wonder and joy in life contagious and persisting.

In 1970, computer science was a burgeoning field, and Joe was restive at Bell Labs. We spent the spring vacation of my second and last year at Columbia traveling to various campuses eager to hire him. I often interviewed for jobs too, a kind of tagalong, because I wasn't at all the hot property my husband was. Even so, in those sexist days, hiring a married couple at the same university was simply not done, ill-advised, nepotism, no matter how far apart their fields were. I heard it again and again. One April day, we succumbed to the beauty of Seattle, and Joe agreed to take a professorship at the University of Washington.

We both loved Seattle. It's a splendid city by anyone's measure, and Joe especially loved how easy it was to hike and ski nearby. Its hills, its abundant waters, fresh and sea, were always beautiful to me, surroundings that suffused me with pleasure, a reminder of the San Francisco Bay Area of my childhood. We made good friends; we hiked Cascade Mountain trails; we skied the slopes. I taught English at Seattle Community College, where many of my students were returning Vietnam vets, an education in itself for me.

But in those days, before Microsoft and Amazon, Seattle was a somnolent place in computer science. Just a year after we'd arrived in Seattle, Carnegie Mellon offered Joe the chairmanship of their computer science department, to begin in the summer of 1971. He couldn't say no. I understood why and concurred completely.

2.

My introduction to Pittsburgh and its superstar computer scientists at Carnegie Mellon had been stealthy. It began when I was an undergraduate, typing course outlines and recommended book lists for Berkeley's business school. Mostly what I noticed then was the recurring name of Herbert A. Simon. Next, as I worked on *Computers and Thought*, I saw how many of its contributors had been at CMU. Finally, the two years I worked at Stanford had filled out the map. Each of those experiences had educated me in just how eminent the CMU people were.

By this time, too, I knew that here, in the heart of the Rust Belt, Allen Newell and Herbert Simon had first conjured AI to life. I knew it, but didn't completely understand it, as I would when I came to write my history of AI, *Machines Who Think*.

At a welcoming party of the faculty, I was so starstruck I stood mute before Herb Simon, the man I'd once secretly accused of being the sole scholar in the skeletal field of business scholarship. I barely had the courage to shake his hand.

Newell intimidated me a little less, although he was nearly as eminent as Simon, and they'd already had a fifteen-year research partnership in cognitive psychology and AI. Along with J. C. (Cliff) Shaw, who was a gifted programmer, Newell and Simon had produced the first-ever working AI program, the Logic Theorist.

I met Raj Reddy that night, too, although not for the first time. When I was at Stanford, Reddy had been a nearly anorexic graduate student at the Stanford Artificial Intelligence Lab, working on speech understanding. One night at a party at John McCarthy's, I met his new bride, Anu, who, in her beautiful sari, sat in round-eyed silence, struck dumb by her sudden airlift from Bangalore to Stanford and its eccentric AI wizards. Now, in Pittsburgh, she was poised and self-confident, a woman with a radiant smile and a droll sense of humor.

3.

Since Michael Joseph in London had published my first two novels, *Familiar Relations* and *Working to the End* (this about a woman scientist in the Big Science milieu of the 1960s), I was welcomed to the faculty of the University of Pittsburgh English Department, headed then by Robert Whitman. Referring to the structure we were in, 42 stories of gothic revival, he said wryly, "You'll soon learn to say 'the Cathedral of Learning' without laughing." It happened that Whitman, a specialist in the literature of the theater, was married to Marina von Neumann Whitman, who taught economics at Pitt. She was the daughter of John von Neumann, whose brilliant work in early computing I'd soon encounter.

My students at Pitt were nearly all the first in their family to go to college, and in that respect, they were a refreshing delight. I began work on a novel called *Three Rivers*, about television news in the seventies, set in Pittsburgh, and in the process, made friends with Pittsburgh journalists.

Joe was engrossed by a department that had been great, but thanks to a sudden exodus of several stars, was now close to subcritical. Allen Newell was his indispensable mentor, as Joe pushed hard to bring the department back to eminence by recruiting new faculty, redesigning the PhD program, and tending to his own research and graduate students, especially H. T. Kung, who had come with him from the University of Washington.

Kung and his new wife were bewildered by a complete absence of Asians in Pittsburgh. They had to drive to Washington, D.C., for Chinese ingredients, a make-do until they visited their Toronto relatives at the Christmas holidays, and could return with the real thing. Much later, at Joe's eightieth birthday symposium, Kung—by then the William Gates Professor of Computer Science at Harvard—recalled how, at Carnegie, they'd fought and yelled at each other as they proved theorems together, snatching the chalk from each other's hands. "You're not Chinese!" Joe once roared in exasperation at this seventy-sixth generation descendant of Confucius. "If you were really Chinese, you'd show filial piety."

My high-spirited husband, who'd once played first board and been captain of the chess team at Bronx High School of Science, sat down one night soon after our Pittsburgh arrival to play the best computer chess program then going, known officially as MacHack, but informally as "Greenblatt," for Richard Greenblatt, its MIT-based designer. Using a king's gambit, Joe checkmated it in about six or seven moves. Without a word, he posted the game's results outside his office door and never challenged it again. (Hans Berliner was the world's correspondence chess champion—which meant he played games by mailing moves written on postcards between distant players— and was working on a chess-playing program for a PhD at CMU. He said Joe's winning strategy had never occurred to him. He modified his own work accordingly.)

4.

For all the good things that were happening in Pittsburgh for Joe and me, I was deeply uncomfortable in this city. I knew no one that first winter and spent my time driving around the city and then out into the countryside. I was shocked by the casual ugliness of early 1970s Pittsburgh—the slag heaps no one bothered to remove; the rusted industrial buildings neither worth using nor tearing down; the negligent lack of grace in the row houses built only for factory workers, mill hands, and coal miners, so who cared? Their inhabitants, coming from European villages shaped by centuries of local culture, villages smoothed at least by time, understood that dismal dwellings and backbreaking labor were the price they paid to escape famine or state-sponsored terror.

Any higher aspirations these immigrants might express focused on the churches—the fine masonry and striking stained glass windows of the Catholic churches, the gold-leafed or lapis-colored onion domes of the Orthodox churches, exotic bits of Byzantium caught by surprise in the hollows of the Allegheny Mountains. Human beings love beauty; crave it. These structures testified to that, places their parishioners could pour all their desire for loveliness, for transcendence, for grace. The churches were a collective human cry for deliverance from daily squalor.

Great American fortunes had been realized in western Pennsylvania—Carnegie, Frick, Mellon, Phipps, Westinghouse, many more—but Pittsburgh was in effect then a colony. Those fortunes had been sucked away to the capitals, New York City and Washington, D.C. Token art filled a few local museums, although a new wing and a whole new museum opened during our time there. Andrew Carnegie had scattered small free libraries about (but he'd done that nationwide, and all honor to him for that).

When World War II steel production made streetlights necessary at noon, a post-war Pittsburgh "renaissance" was declared, and the air was fitfully cleaned up. Otherwise, nothing much had changed in twenty-five years. Downtown was drab, ominous with vacant storefronts. The city fathers imagined that erecting a new sports coliseum, Three Rivers Stadium, would resuscitate this rough, beer-and-a-shot town.

5.

As the 1970s began, Pittsburgh was in the storm path of great changes. The industries that had made the city so enviable (towns all over the United States named themselves after Pittsburgh, hoping to replicate its industrial might) were beginning to die. Strip mining was replacing underground coal mining. Newer, less costly, more nimble steel mills were being built in Brazil and Korea, later in China. The world would eventually take its business there because American management declined to modernize and American labor refused to bend.

This was a gradual process, the slowdown perceived in the early 1970s as only a momentary setback; everything would surely return to normal. Meanwhile, my male students could still get summer jobs as mill hands and make nearly as much money in three months as I made in nine. The acrid smell of steelmaking lingered to pinch your nose as you stepped out the front door in the morning. But death was in the air. In the next forty years, Pittsburgh would lose not only its mining and steel industries, but also forty percent of its population.

The second great change was the information revolution. As the industrial age was dying, the information revolution was being nurtured in the factorylike yellow brick buildings of Carnegie Mellon University, nurtured with a conscious goal that the university could somehow “green” the city, bring it into the new age. But the Pittsburgh where I found myself seemed an unlikely place for such a rebirth. Carnegie Mellon was a verdant island of the future in a vast dead sea of the past.

A third great change in the early 1970s was the rise of feminism. Women had finally stirred themselves for decency, justice, and equality, not just for others, as they had so often in the past, but now for themselves.

Each of these things—the death of the industrial age, the birth of the information age, and the second wave of feminism, would come to affect me deeply, and braid indivisibly in my mind.

I observed and wrote as an outsider, putting the city’s wrenching transitions into words. I taught its sons and daughters. But I felt like a graft that hadn’t taken. No matter how hard I tried, in Pittsburgh, I always would.

From my journal in 1977. *You have to have lived in one of the cultural centers—New York, San Francisco—to understand how deadly intense the life of the mind is in the provinces. Making up for geographical distance from what they consider the epicenter, they attack the great issues with a ferocity that would shock the more blasé citizens of the coasts. This comes across in Gladys's novels* [Gladys Schmitt was a local and much beloved novelist who'd enjoyed national success in the 1940s and 1950s] *and maybe accounts for my own sadness. I feel at a distance from everything too, starved, suffocated, deprived. Yet having lived at the centers, I also know how foolish it is to look enviously that way.*

I wasn't alone. Herb Simon's wife, Dorothea, and Allen Newell's wife, Noël, were both from the San Francisco Bay Area, and we found ourselves strangers in a strange land, making the best of it because our husbands were embarked on such a grand and important adventure. Dorothea Simon was stoic and never complained, but Noël Newell suffered from crippling migraines and wrenched my heart. Later, when the novelist Mark Harris moved to town, and we all became close, it was his wife, Josephine Harris, another who'd spent years in San Francisco, who voiced it unambiguously: this is no place for us.

I was younger, and these older women were a cautionary example to me. In turn, they regarded me with both support and envy. I was a new generation, raising questions about women's roles, no longer willing to accept the utter self-abnegation, the frustrations, that had been most women's lot for millennia. My husband's priorities weren't, by definition, more important than mine. If this city didn't suit me, I didn't have to stay. Locked into the old socializations, the older women felt they must suffer and be silent.

Although I was teaching in the English department at Pitt, I turned to the high excitement that pervaded Carnegie Mellon's computer science department for solace and stimulation. (CMU's business school and the drama department were also peaks, and under Newell and Simon's tutelage, the psychology department would flourish.) Not only were pioneering computer projects underway, but just because of that, interesting visitors came and went.

One afternoon I watched Big Iron being moved into the architecturally brutalist Science Hall. For a certain kind of person, Big Iron—those hulking mainframes of the seventies—equaled Big Power, slightly sexy. That evening I had dinner with a young visiting scientist from Xerox's Palo Alto Research Center, and he told me about his vision for a computer you could carry in your arm, along with a bag of groceries in the other arm. Alan Kay was dear, I thought, but absurd with his Dynabook idea.

Absurd. I write these words on a laptop that was Alan Kay's original idea. My smartphone might let me do that, too, except the keyboard is too awkward. But on that phone I read not only text messages and email, but also the works of Joseph Conrad, Henry James, and Charles Dickens on my daily subway rides.

Part Two: Brains

“However certain our expectation
The moment foreseen may be unexpected
When it arrives.”

—T.S. Eliot, *Murder in the Cathedral*

Machines Who Think Is Conceived; John McCarthy Says Okay

1.

Rebuilding an academic department is time-consuming, so Joe and I began going away summers so he could do research undistracted. We spent a summer in Boulder at the National Center for Atmospheric Research, and in 1973, we were invited to Stanford for the summer. It was a joy to be back on a bike, pedaling under the benevolent Stanford sun, seeing old friends and making new ones.

At the beginning of that stay, I heard from Ed Feigenbaum that John McCarthy, who now had a pilot's license, had made an emergency landing in his small plane at a remote Alaskan spot called, in rough translation, the Pass of Much Caribou Shit. He'd been found and rescued only by chance. I'd finished a novel called *Three Rivers* and was casting about for a new project. What about a novel based on these unusual people I knew in AI?

Then, lying lazily in the shade one afternoon, I thought, why write a novel? Why not write a history? It had to be a cinch—I'd interview a few key people, splice the interviews together, and presto! On July 4, 1973, I tried the idea out on Joe—an intellectual history of AI, told by those who were making that history, told while they were still alive. He loved the idea. Five days later, I tried it out on Ed Feigenbaum, who was very encouraging.

But I began to doubt. Did I have the intellectual wherewithal to do this? Ed grinned. You've got lots of friends who'll help you, and besides, maybe you shouldn't get into very detailed analyses so much as the genealogy of ideas and the personalities involved. The time was right, he went on, and the major figures accessible to me—good friends, in some cases. "I'll get McCarthy on board," Ed said and arranged a lunch at Stanford's Faculty Club for me to get reacquainted with John.

2.

Although artificial intelligence hadn't been a field for very long—John McCarthy had only named it in 1956—McCarthy had been a towering figure from the start. His office was in Polya Hall when I first met him, but he soon moved into the hills behind Stanford to oversee Stanford Artificial Intelligence Laboratory, which he'd started in 1966. SAIL was to become legendary. McCarthy already was.

McCarthy gets credit for a host of key ideas in computing generally, and in AI in particular. One major idea is time-sharing. Computing processes are so much faster than human processes that the computer's time can be shared simultaneously among a group of users without the human users noticing. McCarthy was the first to suggest this publicly—obvious, once put into words—and then designed and helped implement several systems at MIT where he'd been earlier and then at Stanford. But implementation was hard work. Daunting technical problems had to be solved, and, as several people have told me, in the age of 24-hour-turnaround for results (me toting the boxes of punched cards, remember?) the idea was either incomprehensible to many, or bitterly opposed by them.

Time-sharing would change the way people interacted with computers, and how they interacted with each other via the computer. The machines, faster than us even then, were speeding up continuously. Time-sharing is still fundamental to the Internet's operation and underlies everything, like the World Wide Web, cell phones, servers, and cloud computing.

John McCarthy had dreamed up a workable time-sharing system simply because he wanted to pursue his AI research more effectively. In *Machines Who Think,* I wrote about how he and Marvin Minsky organized the embryonic Dartmouth Conference in the summer of 1956, where the first serious practitioners of the art met to talk about the possibilities of intelligent machines.

McCarthy strongly believed that human-level intelligence in a computer might be achieved by using mathematical logic, both as a language for *representing* the knowledge that an intelligent machine should have and as a means for *reasoning* with that knowledge.

In that, his beliefs fell unequivocally into that category of AI known as *thinking rationally*, employing the "laws of thought," embodied in formal logic and mathematics. Knowledge representation was a hazy idea at the time; only later did other AI researchers realize that it was fundamental to intelligent behavior, and, as McCarthy was the first to argue, it needed to be explicit.[1] Knowledge representation

[1] Early 20th century art explicitly queried representation, too. Examples include Picasso and Braque with cubism; Magritte's surrealism (Ceçi n'est pas une pipe); Duchamp's Readymades; or James Joyce's Ulysses. But it wouldn't be until the introduction of

would bedevil and demand refinement in AI for decades: the neats versus the scruffies (the mathematical versus the nonmathematical) until, in the early part of the 21st century, knowledge representation would become one structural element in a tentative bridge between the Two Cultures.

McCarthy's concept, that thinking involves *both* knowledge *and* reasoning, led to his invention of a programming language called Lisp, not the first list-processing language, but certainly the first to be generally useful. List processing is a way of manipulating lists of lists, convenient for representing and acting on attributes of entities, whether objects or actions in, for example, tree-like fashion. An earlier list processing language, IPL-V (Information Processing Language 5) designed by Newell, Shaw, Simon, and Feigenbaum, seemed to McCarthy a great idea, cumbersomely realized. He was proved correct by how widely accepted and long-lived Lisp became. In addition to his work in AI, McCarthy also made fundamental contributions to the mathematical theory of computation.

In 1971, in recognition of all this, he was honored with the Turing Prize, computing's Nobel. He'd go on to win many other international awards and be inducted into both the National Academy of Engineering and the National Academy of Sciences.

3.

But at this 1973 lunch, when I was proposing to write a history of AI, I knew him best as the founder of SAIL. It was McCarthy's brainchild and would be the incubator of some of the most famous

computation into humanities scholarship that those scholars would have reason to query the precision of their own modes of representation. Chapter 26 has more about this topic.

names, techniques, and programs in the history, not just of AI, but of computing generally.

When I'd worked at Stanford, SAIL was a diverting place to visit. Perched on a grassy golden hillside, it had views in clear weather of peaks as far away as Mt. Tamalpais in Marin County, Mt. Diablo in Contra Costa County, and Mt. Hamilton outside San Jose. I'd drive up the winding approach road and smile to see signs warning me to beware of robot cars. An OK Corral moment between my VW Beetle and a mobile robot? I might've stopped to play a game of volleyball from time to time.

Inside the building, the computers, monoliths in the center of a great open room, had been named by the graduate students: Gandalf, Frodo, Bilbo, Gollum, maybe even Sauron. Because we were all taken up with J. R. R. Tolkien's *Lord of the Rings*, everyone's office doors had Middle Earth numbers, rendered by computer in elegant Elvish script, itself a graphics first (not because it was Elvish, but because it was script).

A freestanding robotic arm stood in one area, rather more free than it needed to be, because it was eventually confined behind a Plexiglas barrier to keep it from whacking innocent passersby.

SAIL would be the prototype Silicon Valley organization, with its free-and-easy, 24-hours-a-day atmosphere. But it also laid the foundation of another Silicon Valley tradition. Genius though John McCarthy was—not a word I use lightly—he had little patience for worldly details. He knew enough to hire another gifted scientist as deputy director to make the trains at SAIL run on time. This was Lester Earnest, who was comfortable with the whimsy and the hot tubs, but knew how to set clear goals, get graduate students to meet them, and produce results.

Over the years, results surged forth. SAIL's students and alumni made significant contributions to robotics, to computer miniaturization (laptops and smartphones), to graphical user interfaces (what appears on the screens of those devices), to voice recognition and understanding (Siri, Alexa, and the voices of airline reservation systems or prescription drug ordering). Spell-checkers and inexpensive laser printers came out of SAIL, and video games went from simplicity to complexity. The technical contributions make an even lengthier list.

When I'd first known McCarthy in the mid-1960s, he had an intense sense of social justice, maybe its genesis his politically radical parents. McCarthy taught himself Russian and, beginning in 1965, visited and taught in the former Soviet Union. Later he championed freedom of expression for Soviet dissidents, pressuring the Soviet government to allow them to travel more freely. It's said that for a particular dissident, he illegally brought a fax and a copier into the Soviet Union.

Around 1966, he became active in anti-Vietnam War protests, and one day came into my office at Stanford to ask me to sign a pledge, for publication in the *Stanford Daily*. It declared that the Vietnam War was so reprehensible and contrary to all that the United States stood for, that we signatories would gladly receive, shelter, and otherwise aid any deserters from the armed forces. I believed the war was reprehensible, probably criminal, but I was worried about the risks I might run by making such a pledge. I hesitated.

McCarthy waited, giving off a silent, indisputable righteousness.

I signed.

Just as I was about to go to New York City the first time, my marriage over, I met John McCarthy at a party at Ed Feigenbaum's. He followed me out and invited me for a cup of coffee. We found ourselves in a garishly lighted coffee shop on El Camino Real, all turquoise surfaces and orange light fixtures, and made small talk, which neither of us was very good at. To be with McCarthy for a few moments was to be awed, even made uneasy, by his intensity. He'd had a bone-deep commitment to the austerity and correctness of formal logic. Relaxing his standards to write about AI without using theorems was genuinely painful for him. For one thing, it implied that humans themselves didn't conform to that austere logic. All that made me worry that McCarthy had relaxed his standards just to be sociable with me.

We might have talked about pop music, which we both loved. Sometime just before this, I'd stood in line with a friend at San Francisco's Winterland Ballroom, waiting for the doors to open for a Janis Joplin show. McCarthy wandered past. We chatted for a moment, and I asked him if he'd like to step in front of us into what was getting to be a longer line by the minute. "No," he said politely, "that wouldn't be right." He made his way to the back of the line.

McCarthy went to Czechoslovakia in November 1968, just after the Soviet invasion and repression, and wrote a chatty letter to his deputy, Lester Earnest, giving him not only clear-eyed descriptions of the physical and political results of the invasion and quick evaluations of the scientific groups he visited, but also the music he was hearing—all of it American. When he traveled on to Austria, he wrote, he was delighted to hear West Coast music, which he hadn't heard in Czechoslovakia: Blue Cheer, Jefferson Airplane, plus the Beatles, Otis Redding, and Wilson Pickett.

So maybe he and I talked about music across the table in our El Camino coffee shop, me staring at his *Struwwelpeter* hair, his beard a horticultural wonder. His long dip into the counter-culture much amused him. "Most of the people there have ambitions to put together a 'key,' a kilo of pot, the better to set themselves up in business. They're exactly the capitalists they're railing against," he laughed.

After a while I said, "You are considered—odd."

He gazed at me in silence, and I was sure I'd offended him. Finally he said, "No. I'm just shy."

I had no words for that disarming self-disclosure. I was deeply ashamed that I, who'd suffered two, maybe three, long periods of paralyzing silent shyness in my life, hadn't recognized this.

4.

As John McCarthy met Ed Feigenbaum and me for lunch on a summer midday in July 1973 at the Stanford Faculty Club, he looked superb—all clear-eyed and pink-cheeked, not at all bothered by his recent scrape in the Alaskan wilderness. His hair was still fairly long (but ruly, I noted in my journal) a leftover from his fling with the counterculture. His beard, showing the first streaks of gray, was now trimmed, though it was square-shaped, because he kept tugging on its corners as he spoke.

At first he discouraged me. It was too soon to be writing any histories of AI, he said, the major ideas had not yet emerged. And anyway, why didn't I write about—John had some arcane mathematical project he thought would make a better book. Over the iced tea, I shook my head. "I'm not a woman in search of a project," I said, with more confidence than I felt. "I want to do the history—so far—of artificial intelligence."

Ed excused himself, but John and I spoke for nearly four hours that afternoon. Story after story poured out of him, all amusing, all incisive. In my journal, I noted:

> *It's a pleasure to hear him talk. He brings a wonderful intelligent optimism to life, as if people really can be persuaded to do what's best for them if only it's approached right. He's completely at home with technology, and wonders at the prejudice so many people have against it. If you mention a technological headache, he'll reply firmly, "But there is a technological solution…" Not only is he infectious, but egad, what a delightful change from all the congenitally gloomy people I've been around. Though it drives his colleagues up the wall, I find his sense of play delightful. He seems to be homo ludens in the best sense. Joe tells me that Newell and Simon are often perturbed because two of the elder statesmen in AI, McCarthy and Minsky, refuse to behave in a* "responsible" *manner. I don't know enough about Minsky, but I say, viva John!*

At the end of those delightful, stimulating four hours, the patient staff at the Faculty Club apparently resigned to such marathon conversations, McCarthy said, "Well, mutter mutter, it's your time." "But John," I countered, "you're your own best argument for someone doing this. I'm much more enthusiastic about it now than before we talked."

It was true. Still, McCarthy must never have liked the book. Much later, when he was asked at a grand anniversary celebration of SAIL why computer scientists didn't write their own histories, he said that wasn't their job. But added that the histories that existed weren't very good. I hope he meant only that *Machines Who Think* was by then way out of date.

That summer afternoon, McCarthy shrugged and agreed to cooperate. With Feigenbaum and McCarthy cooperating, and Newell, Simon, and Reddy at Carnegie willing to go along, if only to keep me busy so I exercised no pull on Joe to take him away from Pittsburgh, I needed to get the people at MIT to agree.

I don't know now who made that connection for me, but the connection was made, and I'd go to Cambridge, Massachusetts often to interview them all—Marvin Minsky, Seymour Papert, Ed Fredkin, Joseph Weizenbaum, Joel Moses, and even the reclusive Claude Shannon, the founder of information theory, who'd retired to his fine old Victorian house in Somerville, Massachusetts.

In retrospect, the kindness, generosity, open-heartedness, and candor of each man I interviewed astounds me. Most of them were at a peak of their research careers, in hot pursuit of the next discovery. They had pressing work to do not only for the research they were conducting, the graduate students they were overseeing, the undergraduate classroom teaching they owed their universities, but also for fund-raising, that punishing academic treadmill. Computers were costly; robotic equipment was dear; and the money necessary to pursue AI was, by the standards of the time, colossal. Yet they opened their minds to me for hours, struggling patiently with the elementary questions I asked and with questions nobody had ever asked them before. I was humbled by that. I still am.

5.

Before my eyes, the book laid itself out for me. Western literature has a long rich tradition of imagining intelligence outside the human cranium, beginning with Homer (robotic attendants assist in the forge of the god Hephaestus, who limps badly, they show up as party help, and power the ships of Odysseus around the Aegean); the sages of the European Middle Ages and their brazen heads (both sign and source of their worldly wisdom); mad Paracelsus and his homunculus; Joseph Golem, spying on the gentiles of Prague; Dr. Frankenstein's canonical monster; the robots of *R.U.R*. The endearing and menacing robots of that worldwide phenomenon *Star Wars* appeared while I was writing, and robots have become a staple of TV, movies, and video games. As I did my research, I discovered many early quasiscientific attempts to create intelligence outside the human cranium, so it wasn't just dream-weavers who sang this song.

I'd use that historical framework to make two major points. First, the urge to create intelligence outside the human cranium is an enduring human impulse, with mythical examples across the ages and cultures. Second, this impulse was finally coming to realization in a scientific field called artificial intelligence.

Two attitudes to all this prevailed side by side, I'd say. The first was a general delight in AI, what I called the Hellenic view, because Homer's robots had been welcomed and useful to the Olympians. The other I called the Hebraic view, a fear and sense of sacrilege these creations engendered based on the commandment that graven images were forbidden.[2] That fear and sense of sacrilege moved through literature (and life, as it would turn out) in examples of

[2] For years, I forgot I borrowed these terms from Matthew Arnold's Culture and Anarchy, which I read in my freshman year of college.

such creations gone rogue—Joseph Golem, Frankenstein's nameless monster, the Sorcerer's Apprentice, and on and on.

Such a thrilling narrative must seize my English Department colleagues at the University of Pittsburgh. How could they resist? Literature brought to life. It would captivate any intelligent reader.

And what did I think about AI myself? I was agnostic. I simply couldn't judge the scientific importance of what was underway. But surrounded by the most intellectually exciting human beings I'd ever known, I easily gave them the benefit of the doubt.

The optimism about all this I shared with the AI people was impermeable. We didn't know when, but AI would inevitably come (art is long, life is short, success was very far off). Its arrival would be an excellent thing, for as I've said, pursuing more intelligence was like pursuing more virtue. It had everything to recommend it. Our only adversaries then were the scoffers and the skeptics: *this can never be done*. How I failed to see that this was a human endeavor and would bring in its train all the flaws that have persistently bedeviled so much of human behavior, I can't explain.

I failed to see what else lurked in the shadows.

Perhaps worse, I failed to know my own subconscious wish for AI, what its success might virtuously bring about. That wish was buried and wouldn't surface for another three decades, to be deeply disappointed by AI's flaws even as its influence dramatically ascended in the first two decades of the 21st century.

Over Christmas, We Invented a Thinking Machine

1.

Before I could begin to write, I embarked on a crash course though the AI literature, beginning with *Computers and Thought* (1963) whose substance I'd barely understood when I worked on it. Then I picked up Herbert A. Simon's brilliant little classic, *The Sciences of the Artificial* (1969). Both were lucky choices, for each laid out relatively simple principles that later would be richly elaborated in the field's research.

In *The Sciences of the Artificial,* Simon argued that artificial phenomena—whether a business executive's behavior, the fluctuations of economics systems, or the way people's thinking is influenced by individual psychology, anything that wasn't physics or biology—deserved empirical scientific attention as surely as the natural sciences.[1] "Artificiality is interesting principally when it

[1] This assertion would have gone without saying during the Enlightenment, when thinkers aspired to be "the Newton of the moral sciences." The subsequent Romantic period insisted on a division between the natural sciences and knowledge about humans, or "the humanities." These days, Nobel Prizes are awarded for knowledge about humans, both biological and psychological. Moreover, the sciences of complexity

concerns complex systems that live in complex environments. The topics of artificiality and complexity are inextricably interwoven" (1981).

We live for the most part in a human-made environment, he went on, its most significant element those strings of artifacts called symbols, which we exchange in language, mouth to ear, hand to eye, all a consequence of our collective artifice.

Simon also argued that complex behavior appears in response to a deeply complex environment—an ant finds its way over the ground responding to its environment, not because it has any grand plan to get from here to there. But the ant has a goal and knows through trial and error when it has reached that goal. This image is another version of Simon's longtime preoccupation with the maze, which I'll say more about in the next chapter. That complex behavior appears as a response to a deeply complex environment is a potent idea, and emerges in many different ways in AI.

Two profound principles were explicit here. First, *complexity arises out of simplicity*. Simplicity can be a single neuron that, with its fellow neurons, eventually produces great complexity of thought. Simplicity can be a zero-one state of a computer register, which, together with its fellow zero-one "cells," leads to great complexity of thought. Each example is intelligent behavior, our most human characteristic. Second, our rich human languages are artifacts: we invented them and elaborate upon them daily.

have been the central theme of the independent think tank, the Santa Fe Institute, for more than a quarter century. The term computer science means the study of the phenomena surrounding the digital computer, which is full of surprises every day. But as Simon was writing, these were novel notions.

Computers, Simon continued, are empirical objects, capable of storing symbols, acting on them, copying them, erasing them, comparing them. This happens to match much of what the human nervous system does with symbols, which therefore suggests that some parts of human cognition can be modeled, or simulated, on a computer. Intelligence is the work of a symbol system. It can be enacted in the human brain or in a computer.

Simon's argument foreshadowed an idea that would eventually become a given in the sciences of the mind: intelligence arises not from the medium that embodies it—whether flesh and blood, or electronic components—but from the way interactions among the system's elements are arranged. Sixty or more years later, this idea would come to be called *computational rationality,* encompassing intelligence in brains, minds, and machines.

2.

I needed to go back to the beginning of AI, whenever that was. In the first interview I conducted for my book, Ed Feigenbaum told me a story I recounted verbatim in *Machines Who Think* (1979):

> I was an undergraduate senior, but I was taking a graduate course over in GSIA (the Graduate School of Industrial Administration at Carnegie Tech) from Herb Simon called Mathematical Models in the Social Sciences. It was just after Christmas vacation—January 1956—when Herb Simon came into the classroom and said, "Over Christmas, Allen Newell and I invented a thinking machine." And we all looked blank. We sort of knew what he meant by thinking, but we didn't know. We kind of had an idea of what machines were like. But the words thinking and machine didn't quite fit together, didn't quite make sense. And so we said, "Well, what do you mean by a thinking machine? And in particular, what do you mean by a machine?" In

> response to that, he put down on the table a bunch of IBM 701 manuals and said, "Here, take this home and read it and you'll find out what I mean by a machine." Carnegie Tech didn't have a 701, but RAND did. So we went home and read the manual—I sort of read it straight through like a good novel. And that was my introduction to computers.[2]

Over the Christmas holidays, Allen Newell and Herb Simon had invented a thinking machine. It was one of those moments so modest in the telling that the world would be unaware of the momentous changes to come, like the instant a single-celled creature added a cell or two and became multicellular; or the first hominid stood up on her hind legs, the better to survey the plain; or when some early 20th-century physicists proved to themselves, if not yet to anyone else, that the physical world was not what it seemed.

Speculation about thinking machines, I discovered, had persisted throughout history and cultures—the early Egyptians, the Greeks, the early Chinese, the Japanese. By the 19th century, scientists and poets shared such speculations, especially scientists who longed to create something real and practical. The record says scientists were nearly all driven by no-nonsense goals. For instance, the brilliant English inventor Charles Babbage wanted to calculate by machine (and automate printing—a source of many errors) tables that were essential to navigation and ballistics, tables whose production had,

[2] The IBM 701 was a vacuum tube machine with a total memory of 2048 words (words roughly equivalent to bytes) of 36 bits each. Addition required 12 microsecond cycles; multiplication or division required 456 microsecond cycles. The desktop I write this on has 4 gigabytes of memory, or 4,000,000,000 bytes, and works, well, fast enough for me. Iliano Cervesato, a teaching professor of computer science at Carnegie Mellon, notes that the average smartphone has more computing power than the entire world had in the 1970s, carries out more tasks than were imaginable just a few years ago, and is so cheap that half the earth's population can afford one. Iliano Cervesato, "Thought Piece." Welcome to the <source> of it all. A Symposium on the Fiftieth Anniversary of the Carnegie Mellon Computer Science Department, 2015.

until then, been the task of bored and overeducated clergymen, marooned on the bleak English moors and fens. "We shall substitute brass for brain!" thundered Lord Kelvin, as he oversaw construction of a machine to calculate the elementary constituents of tidal rise and fall (McCorduck, 1979).

But I wonder. When Babbage ran out of money to build his Analytical Engine, he and Ada Lovelace, his associate, thought to raise funds by building a tic-tac-toe or a chess-playing machine. I don't know if that was before or after their proposed system to play the ponies.

These deeply perceptive people must have known their machines offered something well beyond the calculation of tidal tables. If these nineteenth century pioneers left no written record of that insight, there were sensible, even compelling, reasons to keep quiet.

The superiority of machines over muscle was the core of the industrial revolution, transforming life dramatically at the very time these Victorian pioneers were speculating about machine intelligence. Might machines, strong and tireless, excel at thinking, too? Such a fear was—and is—fundamental.

3.

While I was on a crash course through AI's scientific literature in 1973 and 1974, I was also trying to raise money to support my new project. I needed funds for travel and to pay a transcriber of the interview tapes I was making.

I begged from every source I could think of. The Office of Naval Research granted me a meeting, and allowed as how it was an interesting project and certainly should be done. They'd think about it.

The National Science Foundation said essentially the same thing but added with a frown: didn't I realize I wasn't a trained historian of science? Merely a writer? Surely the trained historians of science would be crawling all over the place, I replied. A new scientific field aborning: what could enthrall a science historian more? But I planned to write a book for the general reader, not for other historians of science.

I may have written my first in a series of unsuccessful proposals to the Guggenheim Foundation.

I approached a program in the National Endowment for the Humanities whose explicit purpose was to support projects that married the sciences with the humanities. Here I was, a faculty member in a university English department, ready to write about this entrancing new science, whose genesis lay in some of the dearest and most persistent myths and legends of world culture. The National Endowment for the Humanities couldn't say no fast enough. Its reaction was a harbinger, but I didn't know it. I knew only that I failed to please, and that was that.

And then, as if by magic, money appeared. It seemed to come from Allen Newell, maybe Raj Reddy, but they said no. It was from somebody at MIT with a great interest in the project, who wished to remain anonymous. After *Machines Who Think* was published, I discovered that my anonymous benefactor was Ed Fredkin, then at MIT, whose private foundation sponsored oddball projects and had decided to sponsor mine. I am ever deeply grateful. Without that help, the book wouldn't have been written.

The experience made me skeptical of the whole grant proposal process. No matter who was making the decisions, judgments seemed arbitrary and timid. Many years later, I asked an historian of science, now studying scientists in AI, why I hadn't encountered any of her colleagues at the dawn. "Oh," she said, a little embarrassed, "we weren't sure it would turn out to be important."

So I had some money and was on my way. I only needed to decide how to tell this fabulous contemporary tale. I'd learn much in the course of writing the book, but more important, I enriched my life by getting to know better the geniuses who, with their own intelligence and sense of adventure, had invented thinking machines. Not only were their passports mostly American, but they were in the American grain: optimistic, inventive, pragmatic, plain spoken, and up for fun.

4.

One of them was Herb Simon.

Over Christmas, Allen Newell and I invented a thinking machine.

When I told this story that I'd heard from Ed Feigenbaum to Simon himself, he laughed disbelievingly. "Did I say that?"

In 1971, when we first met, Simon was 55, still brown-haired, though gray was beginning to show at his temples. As a rule, his visage was grim, almost truculent. His face seemed to rest in a distrustful snarl, and his small brown eyes looked out skeptically from under low eyebrows. It was an astonishing masquerade for a man who liked to laugh as much as Simon did, a man who found delighted wonder in everything, whether it was the grackles in the pine trees beside my Pittsburgh house or the endless entertainment of doing science. As I listen to the interviews we had, they're often punctuated by loud laughter, for he was teaching me, and he loved to teach, too. It was also the laughter of two flirts, practicing their art shamelessly on each other.

After the Christmas holidays of 1955-56, Simon, Newell, and their longtime RAND collaborator, Cliff Shaw, didn't yet have a running program on the computer. But they knew how to organize the process, and Simon had recruited his family to enact how it would work—okay, step forward, Dot, and Peter, you fork left over to Barbara. The exercise told them that, with much coding, the program they envisioned, called the Logic Theorist, to prove theorems in Alfred North Whitehead and Bertrand Russell's *Principia Mathematica* could be done. After that watershed December result, Simon claimed it was all over but the shouting.

It wasn't, of course. The coding was arduous and prone to bugs. Many late-night, long-distance, budget-busting phone calls took place among Simon, Newell, and Shaw. Newell said of Shaw, who at the time was in Santa Monica executing the team's ideas on RAND's Johnniac,[3] he's "not just a programmer, but a real computer scientist in some sense that neither Herb nor I were."

In Simon's 1991 autobiography, *Models of My Life,* he wrote that it was 1954 when he and Newell seized the opportunity to use the computer as a general processor for symbols (hence for thoughts) rather than just a speedy engine for arithmetic. Neither of them was ever interested in numerical computing. As I've said, Charles

[3] Johnniac was named for John von Neumann. A copy of his machine at the Institute for Advanced Study in Princeton, the Maniac I, was built at Los Alamos. Maniac was purported to be an acronym for Mathematical Analyzer, Numerical Integrator, and Computer. David E. Shaw's Non-Von was a later computer architecture named for von Neumann.

Babbage and Ada Lovelace saw this possibility in the mid-19th century; Alan Turing and Konrad Zuse had seen this at the end of the 1930s.[4] At the 1948 Hixson lectures at Caltech, John McCarthy was an undergraduate in the audience and heard comparisons between the brain and the computer that determined the rest of his research career: he knew he wanted to design machines that could think. In 1950, Marvin Minsky, still an undergraduate at Harvard but under the influence of psychologist Warren McCulloch and mathematician Walter Pitts at MIT, nurtured the same ambition. Thinking machines were in the air.

With access to one of the best computers of the time, RAND Corporation's Johnniac, Simon and Newell had the means to realize this elusive, longstanding, maybe hubristic ambition: a thinking machine.

[4] Newell said that neither Babbage nor Turing had influenced him and Simon. He didn't know much about either of these early forays into AI. No, what was to be done was just so obvious.

What the First Thinking Machine Thought

1.

What did the first-ever thinking machine think about? Once, this question would have been easy to answer. I'd have said: the first-ever thinking machine was called the Logic Theorist, and it tried to prove theorems in Whitehead and Russell's *Principia Mathematica*.

Now, the question is more difficult to answer because, since the mid-1950s, we've expanded the term *thinking* to cover a much wider range of cognitive behavior.

As I noted in Chapter 2, scientists have been studying cognitive behavior—and calling it intelligence—in nearly any animal (and even plant) imaginable, from whales to amoebas, from octopuses to trees. Biologists are easy with the idea that cognition is ubiquitous, from the cell on up, from the brain on down. This inclusive sense of intelligence was implicit in that first book of readings in artificial intelligence, *Computers and Thought*, where articles about pattern recognition by machine were side by side with articles about chess-playing programs. Computer scientists now call the grand realm of intelligence in brains, minds, and machines *computational rationality*.

But in the mid-1950s of the Logic Theorist, thinking meant only symbolic thinking, the kind of planning, imagining, recollection, and symbol creation that humans alone exhibited. John McCarthy used to tease: maybe a thermostat can be said to think. Was he nuts? No, the proposition was meant to force us to define the specific differences between what humans and a thermostat do. As time has passed, the dividing line is no longer so distinct, and decades of research have shown us that thinking is far more complex than we dreamed.

So let me refine my opening question: what did the first *human-like* thinking machine think about? It was called the Logic Theorist, LT, and it tried to prove theorems in Whitehead and Russell's *Principia Mathematica*. Although its subject was logic, the program was squarely in the category of "thinking humanly," as distinct from "thinking logically." Allen Newell and Herbert Simon were practicing cognitive psychologists and wanted to model the ways humans proved theorems, not create a killer machine that would out-think humans—although each of them conceded that inevitable outcome. They were well aware of other aspects of intelligence, but they aimed to begin by modeling some parts of the highest level of human thinking, the symbolic processes that have begotten culture and civilization.

Given the primitive tools of the time, scientific knowledge about human thinking was scant. Newell and Simon's approach was not stimulus-response, associative memory, or any of the other mid-20th century guesses cognitive psychologists had made about how human thinking worked. LT was a model of human symbolic thinking. It was dynamic, nonmathematical—symbolic—and changed over time.

LT was a first step for Newell and Simon in their ambition to understand the human mind. It would eventually lead to more abstract ideas about intelligence in general.

Details about the LT program can be found elsewhere, but its outstanding characteristics were first that it could learn. Helped along by some heuristics, rules of thumb that the programmers had taught it, the program didn't search every possible path to prove a theorem, but instead considered only likely paths to a proof. As it pursued those paths and met new theorems, it acquired new knowledge, which it stored and then used to solve other problems.

Next, LT could recombine knowledge it already had to create something entirely new. It could quickly widen its search for answers well beyond human capacities for search, which is how it came to discover a shorter and more satisfying proof to Theorem 2.85 than Whitehead and Russell had used. Simon wrote this news to Bertrand Russell, who responded with good humor.[1]

Although LT learned, created new knowledge, found new answers to problems, and knew when it did so, all that was in a limited domain. But this capacity for combinatorial search would serve AI well in the future: When Deep Blue defeated Garry Kasparov, a gasp went up from the human audience because Deep Blue had found a move that had never before been seen in human play.[2] When AlphaGo defeated two Go champions in succession (each one claiming to be the best in

[1] But The Journal of Symbolic Logic declined to publish any article coauthored by a computer program. Moreover, some logicians misunderstood that, as cognitive psychologists, Newell and Simon were eager to simulate human thought processes. These logicians created a faster theorem-proving machine and triumphantly dismissed the Logic Theorist as primitive. So we are; so we humans are.

[2] Joe was in the human audience of this match, having been a gifted chess player in his youth, and told me he thought he was the only one rooting for the machine to win. "But this great program is a human accomplishment," he argued.

the world), it did so by finding a Go move that no one had ever seen before. In combinatorial search lay one aspect of machine creativity.

LT offered no physiological theory of how humans think; it wasn't meant to. But it showed that in the narrow task of proving theorems in logic, human performance could be simulated on a computer in ways that satisfied what cognitive psychologists knew about how the human mind worked. Simon believed that ultimately, a physiological theory of human thinking would be needed. (We still await that in deep detail and keep being surprised by what we do learn.) But instead of researchers trying to jump from the complexities we see in human behavior right down to neuron level, LT represented an intermediate level, one that could obviously be mechanized—Newell and Simon had done it. They called it the information processing level.

Today we might say that LT is the first computer model of what psychologist Daniel Kahneman (2011) calls *slow thinking*, System 2—slow, deliberative, analytical. LT isn't a model of the other kind of thinking, System 1—instinctual, impulsive, sometimes emotional. "Welcome to analytic thinking," Ed Feigenbaum said, after I'd confessed that being around computer scientists was changing how I thought. I didn't then understand that like every other normal human being, I'd think both ways for the rest of my life. At that point in the history of cognitive psychology, the notion of two ways (or more) of thinking was pretty much unknown. Most people believed in an eternal bifurcation: There was thinking. And there was not.

Feigenbaum would later argue that LT's great advantage was combinatorial search—guided by its rules of thumb, it could search a bigger space and find solutions faster than even humans as smart as Alfred North Whitehead and Bertrand Russell. Thus LT exhibited a vital characteristic, sometimes a curse, of AI: in its much greater capacity to search, even guided by rules, we cannot imagine everything that AI will find to do. It will go places we cannot imagine or foresee, with unimaginable and unforeseeable results.

In short, AI will always produce unintended consequences. Similarly, as Amelia Earhart pioneered a round-the-world flight in 1937, she *might* have imagined a global network of commercial flights that would eventually come to pass. But could she foresee that this network would contribute significantly to global warming?

This is a vital truth worth repeating: humans cannot imagine everything possible in a search space. Nor can machines. But a machine can go much further and faster, often with unexpected results.

Another of the deep lessons Simon tried to teach me is also embodied in LT. In scientific modeling, levels of abstraction exist, and the study of each level is useful in itself. Everything material might, at bottom, be physics, but studying chemistry and biology, two higher-level organizations of matter, is still useful. Much later, I'd re-encounter the same idea of levels of abstraction in the work at the Santa Fe Institute, specifically its studies of complex adaptive systems, which begin simply, adapt to the environment dynamically, and end up being complex.

2.

Did LT, with its ability to solve problems and learn from those solutions, take the world by storm? Hardly.

"I guess I thought it was more earth-shaking than most people did," Simon laughed. Then he got more serious. “I was surprised by how few people realized that they were living in a different world now. But that’s a myopic, an egocentric view of it, the inventor’s evaluation.” He was still surprised that, even when we were speaking in 1975, nearly twenty years after LT had debuted, so many didn’t realize how the world had changed with this understanding of what you could do with a computer. “There are still well-educated people who argue seriously about whether computers can think. That indicates they haven’t absorbed the lesson yet.”

Part of the problem was that most people’s exposure to computing then was numerical. Computers might be able to count, but could they deal with other kinds of symbols? As late as 2013, I heard a Harvard professor (in the humanities, true) declare that computers “could only handle numbers.” Didn’t he use email? In some literal, simplistic sense, he was correct—zeros and ones—but computers make a distinction, in George Dyson’s elegant phrase, between numbers that mean things and numbers that do things, and computer systems are hierarchically arranged from the simple level of zeros and ones to a level that can imitate aspects of human thought. After all, Beethoven produced sublime music with a mind whose foundation was on-off nerve cells. All symbols are created and have their being in a physical system. “A physical symbol system has the necessary and sufficient means for general intelligent action,” Newell and Simon wrote (1976). Finally, by inventing a computer program that could think non-numerically, Newell and Simon declared they’d solved the mind-body problem. Or rather, it had simply gone away.

Humans and computers were two instances of physical systems that could manipulate symbols, and therefore exhibit some qualities of mind. LT was an example of a system made of matter that exhibited properties of mind.

What, then, is mind? It's a physical system that can store the contents of memory in symbols, Newell and Simon declared. Symbols are objects that have access to meaning—designations, denotations, information a symbol might have about a concept, such as a pen, brotherhood, or quality. The physical symbol system, whether brain or computer, can act upon those symbols appropriately. We've subsequently learned that, as we think, we process not just memory, but also internal and external information from the environment.

The physical symbol system, simply stated but profound, would undergird AI for decades to come. It would seep into biology as a way of explaining how biological systems functioned. It would come to be seen as an essential condition for intelligent action of any generality, always physically embodied. Some in AI research would break away from this scheme, believing that fast reactions to the environment in real time are more important than a fancy internal representation—a mind—but that was much later.

Understanding, Simon argued, is a relation among three elements: a system, one or more bodies of knowledge, and a set of tasks the system is expected to perform.[3] It follows that consciousness is an information processing system that stores some of the contents of its short-term memory at a particular time, aware not only of some things external to it, but of some internal things too, which it can report on. "That's a small but fairly important subset of what's going on in mind," Simon added to me.

[3] This barebones description of understanding led to philosopher John Searle's attack on AI by means of the Chinese Room Argument, which I have more to say about later.

Simon and Newell weren't claiming to explain or simulate all of thinking—only "a small but fairly important subset." This befuddled the bifurcationists. What was a subset of thinking? For them, an entity was thinking or it wasn't. To abstract only certain aspects of cognition and simulate them on a computer didn't make sense to them. Yet any humanist understood synechdoche, where a part of something stands for the whole of something: "Give me a hand." "Boots on the ground." A synechdoche isn't exactly the same as an abstraction of some aspect of intelligence, but such a comparison might have opened a path for outsiders to push beyond all-or-nothing dogmas about thinking and to begin to understand AI.

The bifucationists weren't the only ones who didn't like this approach. Among dissenters were neural net scientists, who hoped to build intelligence in machines from the neuron up. Simon was fine with this. "We're going from A to B. They're going from Z to Y. Our way suits us better, because Allen and I have come out of the behavioral sciences, economics, and operations research, and we know those fields haven't been able to reduce much of human behavior to formulation."

This would change over the decades.

3.

The contents of *Computers and Thought*, that ur-textbook of AI, were divided into two major parts. One was the simulation of human cognition, which contained papers that simulated human problem-solving behavior, verbal learning, concept formation, decision-making under uncertainty, and social behavior. The other major part was artificial intelligence, which, without reference to human behavior, contained papers about programs that recognized visual patterns, that proved mathematical theorems (logic and geometry), and that played games (chess and checkers), and that could manage some early understanding of natural language. If this now seems a bit muddled—we know human cognition uses pattern recognition and many other seemingly mechanical tricks—it represented the provisional understanding then of what was thinking and what wasn't.

In the 1960s and 1970s, nearly all these efforts would break away into independent fields, with their own social structures, journals, specialized meetings, and peer-review groups. Robotics didn't talk to machine learning, and natural language processing didn't talk to constraint analysis. This approach made complete sense socially even though it was nonsense scientifically: intelligent behavior requires an ensemble of skills. The divisions remain as I write, but certain groups are beginning to explore what they can learn from each other and how combining certain subfields can accelerate cognitive computing.

4.

After Simon's thinking machine announcement, the winter and spring of 1956 were deeply productive. Did he have a sense they were doing something momentous, especially since he seemed to have kept every document possible from that time? His files were crammed with notes of possible paths to pursue, ideas to expand, and only time kept him from doing it. He laughed with glee. "Oh, yeah! Oh, yeah! Yep! It seemed obvious."

As their ideas about a scientific definition of intelligence became clearer and more nuanced, intelligence came to describe a reciprocal relationship between an individual and the surrounding culture, a culture built over many generations. (Remember that walk under the eucalyptus trees at Mills College where that insight was given to me?)

Real-world problems are deeply complex, resources of time and knowledge are limited, and the best way to reach a goal requires identifying ideal actions and an ability to approximate those ideal actions, a kind of built-in statistical procedure humans use to allocate time and other resources. “Sometimes intelligence comes most in knowing how to best allocate these scarce resources,” Samuel Gershman and his colleagues write (Gershman et al., 2015). As intelligent agents, tradeoffs are forced upon us, and we better be good about figuring them out.

Newell and Simon saw that humans used heuristics to identify an ideal action, approximate it, allocate resources, and evaluate tradeoffs. Although these informal rules of thumb didn’t work every time, they cut the search space to reach a goal so that humans can cope with the world in reasonable time frames. Newell and Simon’s AI programs and those of their followers relied on heuristics. But in the 1980s, more formal statistical methods would largely replace heuristics in AI, which, together with other methods (and dramatically improved technology) would take AI out of its cradle and into a lusty infancy.

5.

That summer of 1956, Newell and Simon were invited—as an afterthought—to what was to become, over the years, a legendary summer conference at Dartmouth. Its organizers were two other young scientists, John McCarthy, on the mathematics faculty at Dartmouth, and Marvin Minsky, at Harvard. The conference was under the aegis of Claude Shannon, the father of information theory, then at Bell Labs, and it meant to explore the topic of machines that could think at human levels of intelligence—artificial intelligence.

All the invitees had ideas about how AI *might* be achieved—physiology, formal logic—but Newell and Simon arrived with the Logic Theorist, an actual working program.

"Allen and I didn't like the name artificial intelligence at all," Simon said later. "We thought of a long list of terms to describe what we were doing and settled on complex information processing." Which went nowhere. No wonder. All poetry is sacrificed for dreary precision. Instead, artificial intelligence stuck. I'm partial to it myself.[4]

In September 1956, just after the summer Dartmouth conference, the Institute for Electrical and Electronics Engineers (IEEE) had a larger meeting in Cambridge, Massachusetts. Newell and Simon would be sharing a platform with a small group of others who'd come to Dartmouth, including Minsky and McCarthy. That arrangement made it seem as if they were equivalent, when in fact only Newell

[4] Nomenclature has vexed the field from time to time. These days, you'll hear terms like machine learning, cognitive computing, smart software, and computational intelligence used to refer to computers doing something that, in Ed Feigenbaum's old formulation, we'd consider intelligent behavior if humans did it. Sometimes the new phrase has arisen to dissociate from the largely media-produced reputation AI earned as nothing but failed promises sometime in the 1980s, a time when people began to talk about the "AI winter." Usually, the scientists were careful (I think of John McCarthy's caution: "Artificial intelligence might arrive in four or four hundred years"). But not Herbert Simon. In my book Machines Who Think (1979), he explains his reasoning about the four predictions he made in 1958 that didn't soon come to pass. Later I came across his 1965 prediction that in twenty years, any work humans could do now would then be done by machine. No, not twenty years later; not fifty years later. Journalists and other eager promoters, such as the sellers of IPOs, also overreached. But sometimes the abundant nomenclature also reflects the fissuring of fields into subfields. In the second decade of the 21st century, the term artificial intelligence seems to have regained respectability.

and Simon had a produced a working program; the others were still at the idea stage.

John McCarthy thought that he'd report to the IEEE meeting about the just-concluded conference and describe Newell and Simon's work. The two Pittsburghers objected strenuously: they'd report their own work, thanks. Simon remembered some tough negotiations with Walter Rosenblith, chair of the session, who walked Newell and Simon around the MIT campus for an hour or more just before the meeting. They finally agreed that McCarthy would give a general presentation of the Dartmouth Conference work, and then Newell would talk about his and Simon's work in particular. Newell and Simon were first. They wanted—and deserved—the credit.

The two successful scientists would race ahead, applying their techniques to a more ambitious program, the General Problem Solver, which they hoped would solve problems in general. GPS did indeed codify a number of problem-solving techniques that humans regularly employ. But this emphasis on resoning would mislead AI researchers. Reasoning was necessary to intelligent behavior, but hardly sufficient.

Herbert Simon

The Maze as Metaphor

1.

The maze, the labyrinth, drew Herbert Simon irresistibly: a metaphor for understanding human choice and the patterns of a human life. His witty autobiography, *Models of My Life* (1991) was under the unacknowledged influence of an American classic, *The Education of Henry Adams.* But Henry Adams was a man suspended in puzzlement between two worlds, old and new, whereas Herb Simon was an active begetter of a very new world.

As we think, as we live, we choose step-by-step, turn-by-turn, Simon writes. Our environment offers those turns, those choices. Having chosen, we can't go back. What lies ahead, we don't know. We entertain goals, but they're often vague. "They do not guide the search so much as emerge from it," he writes (1991). No single turn we make guarantees to bring us closer to those goals. In such a labyrinth, minotaurs may lurk. We hope not but must press on regardless (Simon, 1991).

Simon was born and raised in Milwaukee—he declared a great book could be written about the disproportionate influence Midwesterners had in shaping 20th century America—and was educated at the University of Chicago, where he received his BA in 1936 and his PhD in 1943. He headed a research group at the University of California between 1939 and 1942 and then taught at Illinois Institute of Technology until 1949, when he moved to Carnegie Mellon, where he remained the rest of his life.

From the beginning, Simon's interests were capacious: municipal administration, political science, mathematical economics, cognitive psychology, computer science. Yet they circled a central theme: how do people make rational choices?

His first book, *Administrative Behavior,* was based on his doctoral dissertation and examined the prevailing idea in business and economics that any completely rational choice must take into consideration all the alternatives, the possible consequences of each alternative, and compare the accuracy and efficiency of each consequence.

No, he scoffed. For humans, complete rationality is absurd. Human rationality is bound by constraints, such as time, available information, where the decision-maker stands in an organization, and so on. Humans practice only *bounded rationality*, as he called this principle. His argument was squarely opposed to that of classical economists, who kept pretending that humans always made omniscient rational economic choices. His dogged belief in the principle of bounded rationality would earn him a Nobel Memorial Prize in Economic Sciences nearly forty years later.[1]

[1] In subsequent years, two more Nobel Memorial Prizes in Economic Sciences were awarded to researchers who pushed the idea of behavioral economics further: Daniel Kahneman (2002) and Richard Thaler (2017). Arguably, the 2018 laureates, Paul Romer and William Nordhaus, also represent behavioral economics, though indirectly.

Simon saw that human decision-making and problem-solving—rational but fallible—were complex and poorly understood. By 1954, he thought the best way to understand these processes was to simulate them on a computer. In the younger Allen Newell—whom he met at RAND Corporation in Santa Monica and brought back to Carnegie Mellon as his PhD student and later a faculty colleague—he found a research partner who shared those convictions. With the computer, they were confident they could model some small but important parts of human thinking.

In the pivotal year of 1956, even as Simon and Newell were absorbed with the Logic Theorist, Simon published a scholarly paper, "Rational Choice and the Structure of Behavior." It proposed ideas that impelled him to write his only short story, "The Apple." A young man called Hugo lives in a castle with many rooms, each room with several doorways to choose from. Though he can move forward through one of those doorways to an adjoining room, he can never go back. He can see into adjoining rooms, but he can never know what he'll find far ahead. As Hugo (you go, Everyman) develops preferences—for food, for art on the walls—the story slowly reveals the cost of pursuing those preferences.

Simon thought the story was finally about the burden of choice, the search for meaning.[2] It can also stand for the artist finding a way through choices to be made, a world where overpainting, rewriting, and erasing are forbidden. Nor does Hugo encounter a beloved (or

[2] "The Apple," whose structure would be familiar to any videogame player, also anticipates the idea of "expanding into the adjacent possible," the mechanism of biological and social evolution proposed by theoretical biologist Stuart Kauffman in Investigations: The Nature of Autonomous Agents and the Worlds They Mutually Create. Kauffman, however, despises the concept of artificial intelligence, no matter how hard I try to educate him.

despised) Other in any of those rooms. Though it's only a fable, Hugo's lifelong isolation is chilling.

In late 1960 or early 1961, when Simon, on a year's leave from Carnegie Mellon, was at RAND, Ed Feigenbaum brought him a copy of Jorge Luis Borges' *Ficciones*. Simon was thrilled to discover "The Library of Babel," a tale of one of the greatest labyrinths. Ten years later, as a visiting lecturer in Argentina, Simon asked to meet Borges. In his autobiography, he recounts that meeting in some detail, how much they found to talk about—philosophy, mathematics, poetry—yet how he finally understood that, although the labyrinth is fundamental in Borges' fiction, no great abstract model lies beneath. "He wrote stories; he did not instantiate models. He was a teller of tales," Simon writes, reconciled to the difference between impulses that drive scientists and those that drive artists (1991).

2.

When Herb Simon told me some of this in 1974, we'd already become friends. After work each evening, he walked home past our house on Northumberland Street in Pittsburgh. (He had for years and would for years more after Joe and I left that house.) I'd just be putting the cover on my typewriter and might catch the top of his black beret or his wintertime *chu'ulla* sailing past the front hedge. I'd lean out the door, ask him if he'd like to stop for a sherry. He nearly always would. These encounters took place almost weekly.

Over sherry, we'd talk AI shop but also range broadly. Simon was interested in everything. In our formal interviews for *Machines Who Think*, he'd say offhandedly, "Oh yes, I used that because I was teaching myself Greek at the time;" or, "that came about because I was teaching myself Hebrew;" or, "by then I'd got hold of de Groot's book, and was translating it from the Dutch." (Adriaan de Groot's study of how chess masters played, as distinct from ordinary players, would help Simon form his theories of how experts became so, how long it took, and how they operated differently from nonexperts, ideas that animated a later best-seller by Malcolm Gladwell.) Simon had such a good feel for languages that he'd simply open a book in the language he wanted to learn and keep reading until he got it. Eventually he could read professional papers in twenty languages and literature for pleasure in half a dozen.

So in our casual conversations, we liked to talk about cognates and how they changed from language to language. We talked about how languages were structured, verb before subject, verb after subject. Once he'd taken up painting so intensively that he had to step back, afraid it would encroach on his research time, but he often carried a sketchbook—so we talked about art. He was a decent amateur pianist, and we talked about music. Late into our sherry, Simon wasn't above doing a little dishing about his colleagues, perceptive and funny. He didn't suffer fools gladly, and he taught me how to look out for them. We talked about everything from the then sterile state of the humanities[3] to the backward state of the board of trustees of Carnegie Mellon, of which Simon was a member. After one such three-and-a-half hour chat, I was sorry when he had to leave and regretted I hadn't asked him one impertinent question: what's it like to be so much smarter than everyone else you meet? Is it a drag? Fun? Don't

[3] At the time, French postmodernism and other word-salad nonsense began to capture an exhausted field. The physicist Murray Gell-Mann swore to me that in his archives is a note Simon passed to him during a meeting of the President's Science Advisory Committee (an body of eminent scientists that would be swiftly disbanded when it failed to support President Richard Nixon's pet project, the supersonic transport plane). According to Gell-Mann, Simon's note says: "Help me stamp out the humanities."

care? But I'd posed enough for one day, including why he stayed at CMU when more glamorous places were always asking for his hand.

Other people must have wondered because he addressed that question in his autobiography. He conceded how competitive he was, but the competition had to be "both stiff and fair." He'd never believed he had to be at Harvard, Stanford, or MIT to win the academic game. He wanted "to win without conspicuous social support, whether from family or university. Then it would be clear that I had won 'fairly,' and not just by using the hidden, or not so hidden, weapon of a superior environment" (Simon, 1991).

As to the first question I didn't ask, how did it feel to be so much smarter, I can only guess that because he was enthusiastically sociable all his life, he didn't mind the Mendelian shuffle that had endowed him so generously and the rest of us with less.

3.

For the history of AI I was writing, I was reading the field's technical papers, forcing myself to understand them. Sometimes the going was so hard I read through tears of bewilderment. But my agnosticism about AI was giving way.

From my journal, November 3, 1974:

> *Herb declares I'm not a real humanist, since I'm willing to admit human values could be transmitted by extra-human forms, e.g., computers. I've come a long way from the time when I took offense at the idea of computers writing novels. Now I think I'd welcome a new form of intelligence to live in parallel with us. To replace us? That's Herb's referendum: we vote to see whether computers ("beasties" he calls them) which are less prone to human frailty, and which share human values, and can perpetuate those values better than ordinary flesh*

> *and blood humans, should be allowed to replace us. I didn't reject it out of hand, which astonished him.*

Although I didn't know it then, Simon had tried out this idea before—he'd once nearly been ejected from a Yale dinner party for even proposing the question. He would've agreed that it was misleading to ask what might perpetuate human values better than frail humans, without conceding, first, that human values were elastic and, second, that any such beasties might have their own values.

One afternoon we talked about national senses of humor. "You're English," he said. "Maybe you can explain the curate's egg to me. I just don't get it." I'd never heard of the curate's egg, so he told me. It derives from an 1895 *Punch* cartoon, where a young curate is having breakfast with his bishop. The caption reads: "Bishop: 'I'm afraid you've got a bad egg, Mr Jones.' Curate: 'Oh, no, my Lord, I assure you that parts of it are excellent!'"

I was convulsed. I couldn't explain why, but I was. Much later I decided that this was the reduction to the absurd of all my proper English parents' admonitions: "Don't make a fuss. Make the best of it. Always be polite. If you can't say something nice, say nothing. No matter what." But that was later. Herb shook his head in puzzlement and left no wiser than he'd arrived.

February 3, 1974:

> *Herb Simon said tonight, quoting St. Augustine, that thought was pure form, an idea I have to mull over, though it sounds plausible. A lively lecture to a packed hall. A delightfully sensual talk, too: "Our intelligence makes sex the set of rich fantasies that it is;" or, "What if you have to pull a little fact out of your head, like the Latin for 'I love,'* Amo.*" "Is he signaling someone in the*

> *audience?" Joe murmured, to tease me, because he knows I have a sweet tooth for Herb.*

April 13, 1976:

> *Then home, and over dropped Herb for tea. A lovely talk about a million topics, most particularly, did he feel let down after he'd given a talk? Absolutely, even after all these years. What makes you good is the adrenaline you pump into yourself for the occasion, and it seems your body can't let you down gradually but slumps instantly. "You're so grateful for crumbs of praise," he added, and I agreed, though wondered, even he…?*

May 11, 1976*:*

> *Reading* Swann's Way *again. Like rich food, one must do it in small portions. But oh! How did I read it eighteen years ago, enchanted by it, when I'd never thought of writing myself? It's such a writer's book. Yet much beloved by Herb Simon, who's read all volumes through twice in the original French.*

For Simon, Proust was the exquisite artistic representation of memory, which had preoccupied him for a long time. But reading, he said in his autobiography, was more than a mere hobby; it was one of life's main occupations. "As with eating, so with reading. I am nearly omnivorous. But my stomach for words is hardier than my stomach for rich foods, so I do not ration myself" (1991).

September 22, 1976*:*

> *A long talk with a student who reads my published words back to me very flatteringly. But as I said to Herb, who stopped for a brief afternoon sherry, we do it, writing, for love. And Herb said, yes, I do science for love too, and we whooped and giggled about our vulnerability—God bless him for being so honest. And laughed too at how we participate in that stylized game: we say in print "Love me, for here I am," and then the critics say, "I don't love you,"*

> *also in print for everyone to see. We could laugh and recognize our own idiocy. Herb is thinking over doing his memoirs, and is dubious. "Writers do it so well because they're writers. Mark's autobiography was so fine it really discouraged me."*

Our friend, novelist Mark Harris, had just published an autobiography*, The Best Father Ever Invented.*

4.

Over those sherries, the Squirrel Hill Sages were born. I was complaining to Simon that my students had the really interesting conversations—the meaning of life, and all—whereas in the University of Pittsburgh's English Department, I was stuck with squabbles over the photocopying budget, whether the course in Romantic poetry should be one semester or two, and other chutes to tedium. Simon said that he and Dorothea, his wife, had run a little salon when they were at the University of Chicago. People met every Sunday night with the understanding that the gathering wasn't for small talk, but for tackling the important themes.

"You could do that here," he said. "But it would be wise to choose a topic ahead of time." Over the summer of 1976, Simon and I corresponded (by U.S. mail in those days) refining the idea: there would be eight of us only; we'd choose a topic ahead of time; we'd meet after dinner, so nobody had to rush around hosting; and if it worked out the first time, once a month seemed fine.

For the first few months, I privately called the group the Squirrel Hill Sages, after Squirrel Hill, the Pittsburgh neighborhood where we all lived. I soon confessed to my fellow Sages, and they laughed and adopted the name. We were Herb Simon and his wife, Dorothea; Allen Newell, and his wife Noël; Mark Harris, the novelist, and his journalist wife, Josephine; Joe and I. None of us was exactly shy, and the conversations were deliciously lively.

Two topics kept us going long after our usual 10 p.m. ending time. The first was *arete*, the Greek notion of excellence.

November 14, 1976:

> *Last night was the second meeting of the Squirrel Hill Sages. Our topic was* arête*, Excellence, and it was a big hit—we yakked for three and a half hours without a break. We decided that* arête *was both private and public, that you must know yourself to have it (hence no false modesty) and you must demand an acknowledgment from others. You must be excellent in comparison to others of the species (thus relative rankings of* arête*). We decided it implied for us specialization, though that might not be what the Greeks intended. A funny insight: Allen Newell isn't at all turned on by the winner of the decathlon, but loves the idea of the other winners of single events—save the pole vaulters, "because that's only a matter of better technology!" Joe and Herb and I howled; was this Allen Newell, a man whose whole career is based on a better technology? As Joe said, he didn't even look contrite. The only people, we decided, who are worse than those who have arête and don't know it are those who don't have it and think they do.*
>
> *The talkers were mostly the men (Noël had already pointed out what a competitive, masculine notion* arête *was). The scholars among us were dazzling—Dorothea noting that St. Paul had used the word three times in one of his letters, variously translated as triumph; virtue or excellence; and knowledge. Allen noted its root suggested warlike valor, and also quoted the* Meno*, which Herb then objected was middle Greek sort of thought, unlike the early phase of* arête*, which intended war-like virtues. Josephine compared* arête *to the Japanese concept of* shibui*, saying that the economy and restraint inherent in or implied by* shibui *is different from the ostentation—or at least lack of humility—*arête *implies. And so it went. Me, I was astonished that*

> *everyone had taken it all so seriously that they'd do homework. But pleased too, of course. Much fun, much fun. But how exhausted I was at the end of it; what concentration it takes to toe dance with the Sages.*

The second topic to keep us late was Criticism, a discussion that took place in Mark and Josephine Harris's living room. Nearly everyone in the room had put something out in the world and watched the critics gorge themselves.

We spoke of doing our work for love, for future love, for present gratification; of the differences between art and science (although art is for laypersons and science for specialists, in places the two overlap). Of the difference in caliber of critics—critics were often nonpractitioners in the arts, quasipractitioners in the sciences; of different sets of standards, slightly more uniform in the sciences than the arts.

The discussion got personal. I watched as Simon's face, usually guarded, displayed a spectrum of expressions. Finally he spoke. "There are two definitions of criticism. One is where the critic looks at your work, explains it or describes it to people who wouldn't otherwise know about it, and if judgment is made, it's tentative and with some respect for the effort that has gone into the work under discussion." (That's how he spoke, in language precise and informed.) "Then there's the *other* kind." He looked around the room, a grin that might be a snarl at any moment. "This is when the critic is only interested in advancing an agenda, and your work is the innocent victim." Ah yes, we all knew.

After a while, I said to Simon: "Who do you write for?"

"Hmmmm."

“Come on,” I teased.

“They’re not all alive.”

“Come on,” I coaxed.

He gave me a look that was jokey, sheepish, and maybe a little proud. “Oh, Aristotle, for one.”

The Squirrel Hill Sages met regularly for perhaps two years, until Joe and I left Pittsburgh. In one sense, we were the living embodiment of cross-disciplines, feeding each other gladly, a negation of the premise of the Two Cultures. Recall my description of how earnest the life of the mind is outside the cultural capitals. But it was also me, my own yearning to bridge the Two Cultures. I started the Sages. At the turn of the millennium, Joe and I started a salon in Santa Fe, New Mexico, where we then had a second home, and the gathering extended for a week between Christmas and New Year’s to tackle meaty issues. The salon met annually for more than fifteen years. I still belong to informal discussion groups with serious purposes in New York City. *Mea culpa*. But such groups coalesce because so many of us yearn for those bridges.

“Terminally earnest,” a friend once teased me. At least I think he was teasing.

5.

Herb Simon was finally a paradox. He was brilliant, generous, visionary, and deeply engaged with the world, at the same time he was shy and despised smalltalk, which made him seem remote. He loved the arts—that obsession with painting, that omnivorous polyglot reading—as much as he loved the sciences. He had not a shred of false modesty, and to paraphrase Churchill, he had little to be modest about. He was intellectually self-confident enough to admit that the Simons subscribed to the daily newspaper so Dorothea could read the news while he read only the comics (a taste he and I shared, though I read the news too). He dressed like the color-blind man he was, blacks and browns, sometimes together, and aside from his few miles walk each day to and from work, I don't believe he ever indulged in other exercise. (He claimed his ambivalence about owning a car had only been resolved when he and Dorothea became too embarrassed to ask their friends to drive babysitters home.)

But contradictory aspects lurked and drove him. For a man who studied and honored rationality, however bounded, he was deeply passionate (his reputation for having a significant temper; those keen flirtations he confessed to in his autobiography that remained chaste, because he was too fearful of being rejected, and because of his deep loyalty to Dorothea). When critics claimed that AI was failed promises, they might have meant some of the claims Simon made: the four predictions he made in 1957, which weren't soon to be fulfilled (he explained why, not entirely convincingly, for *Machines Who Think*); his flat 1965 assertion that any job a human could do would, in twenty years, be done by a machine (not in twenty years; not in fifty). What impelled him to such farfetched stuff?

He was sharply competitive, with a perplexing belief that, for winning to count, he must come from behind. His psychological needs must have been demanding, needs that left little room for his children, two of whom eventually went off to boarding school. (I speculate. The issue is opaque, and might have been Dorothea's; might have been the children's.) He had a well-honed debating style. I'd watch him tangle with anti-AI people in formal debates, and though I agreed with him, wouldn't willingly have exposed myself to that weed-whacker rhetoric.

People like me, like most of his students, like many of his colleagues, loved and admired him. Manuela Veloso, for many years the Herbert A. Simon Professor of Computer Science at Carnegie Mellon, and a world renowned roboticist, remembers that when she was a young faculty member, struggling for publications, she felt particularly down when a paper she thought was excellent was rejected by a journal. Over lunch with Herb, she confided in him. "Of course that kind of thing never happens to you. I mean, you've got a Nobel Prize, and all." He shook his head. "Of course it happens to me. But Manuela, you've got to be your own evaluator. It's got to come from inside. You know when you're doing something good. Don't be too dejected by lack of recognition, or even elated by the opposite. That's a roller coaster. Keep faith in yourself." Veloso never forgot that crucial advice.

People exposed to his darker sides were less warm. As well as affectionate friends, he had an abundance of bitter lifelong enemies and, I now see, he generated considerable professional jealousy. Nobody needed to be that good at so many things.

With Dorothea Simon, I never got past a certain cool friendliness. In retrospect, it couldn't have been pleasant for her to think of her husband stopping weekly for sherry and talk with a much younger woman, but those conversations were so cerebral that, at the time, it never occurred to me. She'd been a gifted and, from photos, a startlingly beautiful young woman, already doing graduate work at the University of Chicago, when she met and married Herb. She was expected then to give up her own intellectual and professional life to serve his. She mostly did, until their children were grown, and she could go back to school for research in education and learning. The Simons even published a few papers together on those topics. When the Squirrel Hill Sages met, she was penetrating, well-informed, and articulate.

She was noticeably cool to my 1970s feminism, perhaps because it seemed to rebuke everything she'd sacrificed herself for. Yet one night after one of Herb's public lectures, I saw her receive gushing praise from a member of the audience for her husband's talk. She responded graciously. But in her eyes, in her stance, was pain. I wondered if she was regretting what it might have been like to receive that praise on her own behalf.

Allen Newell

Brilliance and Puritanism

1.

I've never known a scientist more singularly driven than Allen Newell. Science, and science alone, drove him. Newell was probably the purest scientist—someone who adored doing science for its own sake alone—that I've ever met. He preferred to work and honored work over anything else. He was subtly—and sometimes not so subtly—dismissive of people who didn't share that value. His scientific, intellectual, and moral stature were such that if you thought there might be more to life than work, you felt slightly shamed by your shabbiness.

At least this was his public persona. In fact he read widely (several times I interrupted him and Noël reading aloud to each other in the evening; once *The Lord of the Rings*) and he was not to be reached during Monday Night Football, especially if his beloved Steelers were playing.

Newell's work in computing was wide and deep. At the end of his life, he declared that his career had been devoted to understanding the human mind. This he'd done in a series of computer programs that, as Herbert Simon put it, “exhibited the very intelligence they explained.”

Newell and Simon produced the first working AI program, the Logic Theorist, described in Chapter 9. They collaborated in designing the early list-processing language called IPL-V. Although it was superseded by John McCarthy’s more elegant Lisp, IPL-V laid out the paradigm of lists of lists for both the representation and the solution of nonmathematical problems by the computer. The Logic Theorist and their next program, General Problem Solver, codified some heuristics that humans used to solve problems, such as means-ends analysis (“Here’s my goal; what’s the best way to reach it?”), backward chaining (“If I’m at the goal, what steps did I need to arrive here?”), and the identification of subgoals that moved the program further toward the main goal. At the time, these two programs succeeded enough to confirm what we’d believed since at least Aristotle: reasoning was the glory of human intelligence.

In 1972, Newell and Simon published *Human Problem Solving*, a massive and highly influential book in cognitive psychology that explored the ways humans solved various kinds of problems—slow thinking—based on what psychologists had evoked from human subjects who spoke aloud while solving problems. For example, a subject trying to guess the next number in a series might say, “No, I won’t choose four, because it’s already come up so many times that I doubt it’s going to come up again.”

For Newell’s work toward a general theory of the mind—the *human* mind, where the computer was a means to embody his theories—he’d receive many prizes and honors, including the American Psychological Association Award for Distinguished Scientific

Contributions, membership in both the National Academy of Sciences and the National Academy of Engineering, and honorary degrees. He delivered the 1987 William James Lectures at Harvard, won the National Medal of Science, and received with Simon the Turing Prize, which is computing's Nobel equivalent. But, as Newell once laughingly said to me, by the time the awards come, the game is really over. It wasn't; it still isn't; but it amused him to say so.

Although understanding the human mind was Newell's great goal, he yielded to several diversions, which produced major science and technology in their own right. With Gordon Bell, a giant in computer design, he worked on computer architecture, and they published an influential book in 1971. He played a major role as advisor to ARPA's speech-recognition program in the early 1970s. With scientists at Xerox Palo Alto Research Center (PARC), he worked on the psychology of human-computer interaction, which led in 1973 to the Xerox Alto machine, which had a graphical user interface, was controlled with a mouse (an inspired idea of Douglas Engelbart), and was a forerunner of many of the personal computing environments that followed (such as the Macintosh). Newell also wrote a series of papers and several books.

In the late 1970s and 1980s, Newell was not only filling out his research toward a unified theory of cognition, but working doggedly on hypertext and computer networking so that Carnegie Mellon became one of the earliest of the wired campuses.[1] His ultimate

[1] The committee Newell chaired to explore this in the late 1980s faced heavy weather. In his typical style, he wanted the meetings to be completely open and recorded on the hyperlinked Zog net he and his students had invented so that the committee's minutes could be searched in a variety of ways. Even at CMU, people asked whether the computer wasn't mere gimmickry. Worse, would it attract only nerds to CMU? Could the day-to-day support be maintained by a school that had only recently installed

work, unfinished at his death, was that unified general model of human cognition, called Soar, which would be the ancestor in many prescient ways of today's major AI programs, such as Google Brain, Nell, and AlphaZero.

2.

Newell was a big man with a round face that broke into an easy grin. His porcelain dome seemed to signify intelligence itself. White sideburns popped out cheekily from beneath the earpieces of his glasses. His arms were so long, he had his shirts custom-made. I can see him shambling down the halls at Carnegie Mellon, a high-school football player's body matured into softer stuff, stopping to talk to graduate students and colleagues. In meetings, he knew how to listen, but also how to get quickly to the heart of the matter. (His two characteristic phrases became common currency in our household: "That's not entirely ridiculous," meaning worth some consideration, and "Are we there?" which meant he thought the decision was arrived at, the problem solved—quickly and usually wisely, in his case.) He was a composer and conductor of the symphony of a profound mind, leading you along new paths, beguiling you with the sheer audacity of his ideas.

Soar was Newell's last big idea and possibly his most audacious. He aimed to construct a scientific theory of mind, detailed and encompassing, with hierarchical layers that were to explain mind from the lowest, the neural level, to the highest, the symbolic manipulation level. That top level included rationality and creativity. The PhD students and postdocs who worked with him on Soar—John E. Laird, now a professor at the University of Michigan

telephones in dormitory rooms? It all went well: the Andrew System of centralized computing and file retrieval was one of the first instances of cloud computing.

and a founder of a firm to commercialize Soar; and Paul Rosenbloom, a professor of computer science at the University of Southern California—went on with their students to develop the model extensively. At a 2014 event called the AI Summit, where AI leaders discussed what the next great steps should be, Kenneth Forbus, a distinguished AI researcher from Northwestern University, named Soar as an example of the kind of grand thinking that once propelled AI research and should again.[2]

Newell was at pains to point out that Soar was a *scientific model*, not mere metaphor. For decades, the computer-brain comparison had been commonplace: "giant brains," early journalists prattled. But Newell wanted to move beyond metaphor to scientific model. He warned, "To view the computer as only metaphor is to keep the mind safe from analysis." Although one must always acknowledge the necessity of approximation and the inevitability of error, a scientific model tries to describe its subject matter directly, not metaphorically.

Newell described this daunting task in his eight William James Lectures at Harvard in 1987. Soar was different from philosophical theories of mind because Newell and his students were trying to instantiate, in executable computer programs, the nature of each level of the model, matching and modifying them to conform with what psychologists then knew about human cognition and the ways one level reacted with adjacent levels, above or below (Newell, 1990). Grand theories of mind are still problematic, but we now know

[2] Three years later, a University of California Berkeley group of engineers and computer scientists issued a report, "A Berkeley View of Systems Challenges for AI," which addressed the kind of cross-disciplinary sharing that future AI systems must incorporate. Stoica, I., Song, D., Popa, R. A., Patterson, D., Mahoney, M.W., Katz, R. H., Joseph, A.D., . . . Abeel, P. (2017, October 16). A Berkeley View of Systems Challenges for AI. (Technical Report No. UCB/EECS-2017-159). Retrieved from http://www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017-159.html

human brains seem to work in a way similar to Newell's proposed multilayer model. Today, when neuroscience and AI are rapidly driving each other to achieve better, more accurate models of human cognition, I regret that Newell didn't live to see it.

Soar was an epic undertaking. But audacity, thinking big, was characteristic of Newell and how he wooed his colleagues. One day, as I was interviewing him for *Machines Who Think*, he speculated that the idea of a physical symbol system—a system that could think embodied in some physical way—was, in its implications, as profound for understanding mind as the idea of natural selection had been for biology. He saw my face light up. "You like that," he declared, didn't ask. Yes, I liked that, as he knew I would.

Yet Newell was aware that to analyze mind by means of a computer would also be to synthesize intelligence in a hardier medium than flesh and blood. Once, in the mid-1970s, he repeated what was common currency then (and now, since I saw it in a 2013 book by Eric Schmidt and Jared Cohen, *The New Digital Age*): in the future, machines would do what they did best, and humans would do what they do best. "But that's bullshit," Newell said, a man not given to coarse language. "The machines will just keep on getting smarter and smarter. There won't be much left for humans to do."

3.

The Newells had deep roots in the San Francisco Bay Area. Newell Avenue in now-urbanized Walnut Creek, across the bay from San Francisco, is named for a cousin of Newell's father, who owned a large orchard in the area, a place where Allen spent many days of his boyhood. Newell's father, Robert, was an eminent professor of radiology at the Stanford Medical School when it was still located in San Francisco, where Allen grew up.

“He was in many respects a complete man,” Newell said of his father. “He’d built a log cabin up in the the Sierra. He could fish, pan for gold, the whole bit. At the same time, he was the complete intellectual….He was extremely idealistic. He used to write poetry. He thought that friendship was so important that he consciously cultivated his friends. He made regular appointments to see them. He actually thought this was important.” Newell said this last with a sense of wonder and just a touch of skepticism.

“And I got my strong sense of ethics from my father,” Newell said with some ambivalence, as if it had cost him pleasures, burdened him uncommonly. He offered the statement to me like a flaming brand, a keen double-edged sword: would I seize it? Come up to his impossible standards? That challenge may have been unconscious, but I felt it nevertheless.

4.

After Allen Newell’s death in 1992, Herb Simon wrote an affectionate and scientifically detailed memoir of Newell for the National Academy of Engineering and mentioned that Newell had met Noël McKenna at Lowell High in San Francisco, where, at age 16, they’d fallen in love. They married when Newell was twenty. “The marriage demonstrated that Allen and Noël were excellent decision-makers even at that early age, for they formed a close and mutually supporting pair throughout the forty-five years of their marriage” (Simon, 1997).

It’s not quite so straightforward. When I could coax Noël off her perennial sickbed, where she'd been felled yet again by chronic migraines, to come and have lunch with me, I'd hear a somewhat different tale.

First, the socially prominent Newells were appalled by Noël McKenna as wife for their young scion. In their view, she was nobody from a nothing family. Shortly before the start of the Great Depression, in Noël's infancy, her mother died, her father deserted the family, and she was thrust upon a struggling widowed aunt to be brought up, considered just another unwelcome mouth to feed. The Newells did everything they could to oppose the marriage. Allen Newell, who was in love and willful, married her anyway. When we met, Noël was ethereally beautiful with perfectly molded cheekbones, large sad brown eyes, and prematurely grey hair pulled back prettily in a low knot. She was delicately built with finely formed hands and a soft but not girlish voice. I saw her once in the gym and marveled aloud to her that she didn't have an ounce of extra flesh on her. "Lovely Noël," I wrote in my journal, "like a delicate fern, beautiful and fragile. Yet in many ways I hardly know her. Is she sensual? Does she have a temper?"

Noël was ridden by self-doubt. In her Dickensian childhood, she was regarded as just another mouth to feed; her aunt often threatened to throw her out on the street, where she'd have nothing, did she understand that? She did. She fled to Allen and his love as the first reliable shelter she'd ever known. She loved Allen for the rest of her life, and Allen did love her deeply—all his life. But he loved science perhaps as much, maybe, she sometimes thought, more. The shelter of his love became for Noël Newell a refuge, but sometimes a kind of prison. With their only child now grown and gone from home, she was alone in an enormous house with a companion whose greatest attention focused on his teletype and the telephone.

Noël and I met from time to time to talk about what we were reading or to see a movie together. We'd both grown up in the San Francisco Bay Area, and our meetings might then devolve into mutual lamentations over being stuck in a gray, ugly, beer-and-a-shot-town because of our husbands' work. We both yearned to be out of 1970s Pittsburgh and back home, the Bay Area. We both worried about the impact on our husbands if we insisted on getting out. "Would Allen lose his muse?" Noël wondered. It was unthinkable to leave, unthinkable to stay.

Late in May 1973, when Joe and I had been in Pittsburgh for about two years, my journal reports a lunch with Noël. She protested feebly that she'd "made her peace with her life as it is," yet added that she was worried that when Allen died, she'd have nothing, not even a roof to call her own—or as she interestingly slipped, "I guess I should be glad to have a roof over my mouth." *When Allen dies.* They were both in their mid-forties, and Newell, at least, was in robust health, except for perpetual back troubles. Although financial worry was irrational—the Newells lived in a grand old house on Marlborough Street in Squirrel Hill that they'd recently moved to because their former house couldn't structurally support the five thousand books in Allen's library—it was a demon of her childhood haunting her yet. Noël was undergoing intensive psychotherapy to treat the migraines, considered in those days solely a psychosomatic disorder, but therapy was dredging up all the other miseries of her childhood, too.

That lunch might have been where I arrived with an idea to cheer her up. Why didn't she think of going around with a tape recorder and asking all these AI people what they thought they were up to? But I couldn't bring myself even to suggest it; she was so down and forlorn that I spent my time with her just listening, sympathizing. Only years later did I recognize with what immense courage Noël lived her life.

She and Allen loved each other deeply—that was clear. But nobody loves profoundly without some difficulties. I myself was having some troubles with Newell. I felt I liked him better than he liked me. I understood that moral challenge, that double-edged sword he'd flung at me. It went like this.

5.

Joe was becoming more prominent in computer science as he simultaneously published significant research and began to return the former luster of the Carnegie Mellon computer science department. (Newell told him much later that the few remaining senior faculty members had agreed among themselves to give him a year: if he couldn't turn the place around, they'd feel free to leave, too.) He not only turned it around; he set it on a firm path to future distinction. Thus Joe stood out at a moment when computer science departments were being formed all over the country and looking for someone to head and build them. He was regularly approached about moving from Carnegie Mellon, but nothing really tempted him until the University of California San Diego called.

California. Warmth and blue skies. My heart leaped. I'd awakened more than one morning and stared out at Pittsburgh's dreariness, with the unwelcome thought that my parents did *not* get me out of Liverpool, England, at enormous effort, for me to end up in Pittsburgh, Pennsylvania. But Allen Newell, with a kind of magisterial contempt that only he could express, proclaimed that anybody who went to California—he called it Lotusland—was "only interested in getting a suntan." For all his geniality (and he was very genial), he had a streak of puritanism, almost self-righteousness, that one hoped not to provoke.

Joe was deeply divided. He too loved the West and the idea of more dimensions to life than research, administration, and teaching. Everybody at CMU worked seven days a week because not only were they immersed in their work, but, especially during the long Pittsburgh winters, what alternatives existed? He didn't like 1970s Pittsburgh. He never learned to find his way around the city. But he deeply loved his department and his work and was very good at it. More than twenty years after we left Pittsburgh, on the occasion of the 25th anniversary of the founding of the computer science department, Catherine Copetas, then an assistant dean, introduced Joe's talk, saying that Joe had "implemented many of the traditions here. Alan Perlis, Allen Newell, and Herbert Simon founded the department, but it was Joe who took this place and turned it into an organizational wonder."[3] He felt immense respect, affection, and loyalty toward his colleagues: a big chunk of his heart would remain at Carnegie Mellon forever.

Joe was caught not only between two kinds of life he might like to lead, but also between two strong people: his wife, who wanted very much to get out of Pittsburgh, and his mentor and deeply admired

[3] In October 2015, at the celebration of the fiftieth anniversary of the creation of CMU's computer science department, CMU's provost announced the creation of the Joseph F. Traub Chair in Computer Science to honor Joe's early leadership. This was nearly 35 years after Joe had left CMU for Columbia. I was in the audience, still stunned and raw only two months after Joe's sudden death. This honor to Joe's memory made me burst into astonished and grateful tears.

friend, Allen Newell, who provided invaluable administrative advice and set a shining example of deepest devotion to the life of the mind.

Probably from the beginning, Newell was wary of me. I wasn't going to be the usual faculty wife, a phrase I detested. I had my own faculty position, unusual in the early 1970s. I'd published two books and was working on a third. I was getting more and more involved with second-wave feminism: I taught a course at the University of Pittsburgh in women's studies, and I was an officer of the Allegheny County branch of the National Women's Political Caucus. I was trouble.

In fact, I wasn't trouble. History was. We were all at an epochal point in relations between the sexes: the easy entitlement any man could once assume was now in question—under assault, some said—and it seemed clear to me then which way history would go. Newell surely felt that I was a bad, even subversive, influence on a fragile woman like Noël.

Late that spring, Joe decided he couldn't bear to leave his department, and said no to San Diego. I was deeply sad. But we too made our peace with Pittsburgh, at least for the time being. We bought a house, and I began the work that would be *Machines Who Think*. Newell exhaled and decided I was okay. I was a serious interviewer, intent on writing a good history of AI, and Newell appreciated that. As the Squirrel Hill Sages got underway, he seemed to grow to like me better, and I was relieved.

6.

When Allen Newell was awarded the U. A. and Helen Whitaker professorship in September 1976 at Carnegie Mellon, there was a great celebratory dinner, and for the occasion, he gave a talk entitled "Fairy Tales," which was to become a classic in computing literature.

Fairy tales, he began, are the way we, as children, learn to cope with the world, enduring trials, overcoming obstacles. But now we are all as children, facing an unknown future. "I see the computer as the enchanted technology. Better, it is the technology of enchantment" (Newell, 1992). Computing is the technology of how to apply knowledge to action to achieve goals. It provides the capability for intelligent behavior, with algorithms that are frozen action, to be thawed when needed. The continuing miniaturization of these physical systems, smaller, faster, more reliable, less energy-demanding, means that everything is happening in the right direction simultaneously. Thus computing offers the possibilities of incorporating intelligent behavior in all the nooks and crannies of our world. "With it, we could build an enchanted land." He went on to say how, but warned that in fairy tales, trials had to be undertaken and dangers overcome. We must grow in wisdom and maturity; we must *earn* our prize.

Over the years, as problems arise with computing in society that are sufficiently grave to make us falter and wonder if we've made a bad trade and might retreat, I've reminded myself: we must *earn* our prize.

Joe and I called Newell the following day to say how good it was. "I didn't see your name on the guest list, so I wasn't sure you'd be there," he said to me. "But then I saw you, and was very aware of you as a professional in the audience, hearing not only what I said, but if it scanned."

It scanned. It, well, soared. Its message suited me perfectly, optimistic but cautionary, weaving together the deepest purposes of story with the promise of a science, and its concomitant technology, that I was falling in love with.

7.

By the time we got to the International Joint Conferences on Artificial Intelligence in Boston in the summer of 1977, Newell and I were pals. We had one long dinner alone together, where he raised an interesting and, for me, evocative theme. "Do you believe in the Two Cultures?" he asked.

I nodded. I knew all too much about it.

"I don't," he said. "I think there's more like seventy-five cultures, none of them able to talk to each other in any sensible way." I must have protested; it had been less than a year since I'd heard "Fairy Tales." Wasn't this a way for those seventy-five cultures to speak to each other? But he wouldn't budge.

We were all spending that summer in California, he and Noël in Palo Alto, where he was consulting at Xerox PARC on human-computer interactions, and Joe and I in Berkeley, Joe with the University of California's computer science department as the guest of Richard Karp. We lived in the Berkeley condominium we'd bought a few years earlier, so I could be near my family and in my beloved Bay Area, at least part of the year. Thus Newell and I flew back to San Francisco together from the Boston meeting, and he was Scheherazade—he entertained me every moment of the more than six-hour flight. We talked about aging ("I'll be glad to lay the burden down," he said cheerfully); about religion; again about the sciences versus the humanities, a topic he'd spent some time thinking about.

He retold Frank Stockton's old story, "The Lady, or the Tiger?" which he loved (Stockton, 1895).

A humble young man and princess fall in love. When the king discovers their love, the man must undergo an ordeal of judgment in the arena: he must choose between two identical wooden doors. Behind one is a beautiful woman, whom he can marry and live a long and happy life. Behind the other is a hungry tiger, certain death. The princess has managed to discover which door is which and promises to signal her lover. She also discovers that the lady waiting behind one door is her rival in beauty and charm. At the moment of trial, she signals her lover subtly. The story ends with the question: which door does she send him to?[4]

Newell thought that this story encapsulated the idea that, given the complexity of the real world, there is no way to predict with accuracy the outcome of a determinate process of any complexity. The computer "does only what you tell it to do," but we can't know exactly what that will be.

We had high serious conversation; we had lowdown gossip. He gave me a lecture (nay, sermon) on commitment: "I get so angry at people who get divorces—and usually very hostile to the guy, because he takes his 75 friends and she takes her 4 friends, and they split…."

So between us finally, all was calm, all was bright. On January 25, 1978, my journal says, "*Allen wrote me a net message* [one of our early terms for email] *and invited me to their house to play with ZOG*." A few days later, I was with the Newells, playing with Zog, an early hypertext system that Newell and his students had developed as a way

[4] Newell includes "The Lady, or the Tiger?" in his address as the first president of the Association for the Advancement of Artificial Intelligence (Newell, 1981).

of accessing psychology and AI programs developed at CMU. Zog was fun, and I thought what a superb writer's notebook it would make. But no writer I knew could afford the hardware, much less the software of such a thing. (True: Zog was implemented on the USS *Carl Vinson* to access its administrative database.) Yet as Newell had said in "Fairy Tales," this technology gets cheaper and better in every way—even impoverished writers now have such things at their fingertips, pretty much for free: the World Wide Web, Wikipedia, not to mention cheap, special-purpose programs for organizing large bodies of prose.

In 1980, as the founding president of the American Association for Artificial Intelligence[5], Newell addressed the newly formed association with a demanding talk called "The Knowledge Level" (Newell, 1981). A precursor to his William James Lectures, the talk proposed multiple levels of cognition that the brain has since been shown to exhibit and that one sophisticated ML technique now employs, in what's known as deep learning. Newell defined the top level: "Rational agents can be described and analyzed at an abstract level defined by the knowledge they possess rather than the programs they run." This is the knowledge level.

This top level is what the system knows about the world in which it operates and can use to reach its goals, including the ability to identify and search for missing knowledge. Humans tend to find and store these search results for future use, although some machines are fast enough to search problem spaces whenever they need the knowledge, without necessarily storing it for the future. From lower levels of knowledge, the system aggregates knowledge at a higher

[5] Today, this organization is named Association for the Advancement of Artificial Intelligence (AAAI).

level (Newell, 1981). This was mostly speculative on Newell's part, but the current work of brain scientists points in the same direction.

Many details could still not be filled in or verified when, five years later, Newell delivered his William James Lectures at Harvard in 1987. (He confessed his personal embarrassment that he himself had not done every experiment, something he'd always vowed to do to verify his scientific theories.) Moreover, in 1987 it would have been difficult to envision that top level, the knowledge level, with the kind of access to vast worldwide data that programs like Google Brain, Nell, or others now have. That said, he seems to be generally correct in his notions of how thinking takes place in the human brain, at multiple and asynchronous levels. This strikes me as the kind of insight only a computer scientist could have had.

Soar, the project in which Newell instantiated the ideas in "The Knowledge Level," and his William James Lectures, showed how a relatively few elements of architecture can combine to produce new capabilities, without necessarily building a new module for each new capability. John E. Laird and others would take Soar further, seeking cognitive Newton's laws, a small set of very general mechanisms that give rise to the richness of intelligent behavior in a complex world.[6]

[6] In Newell's 1987 William James Lectures at Harvard, he compared his proposed (and then, only partial) computer model to what was then known about human cognition, from the lowest, the device level (cellular), to the highest, the knowledge level (the agent with goals, actions, and body that operates in the medium of knowledge—what it already knows from experience, what the outside world provides, employing all these layers to exhibit intelligent behavior). Yes, the two sets of levels, human and machine, differ physically, from electrons and magnetic domains at the device level to (several layers up) symbolic expressions at the symbolic level. But functionally, they were the same. "System characteristics change from continuous to discrete processing, from parallel to serial operation, and so on" (Newell, 1990). The sets of levels also operated asynchronously, some quickly, others more slowly. He boldly proposed Soar as a unified theory of cognition.Significantly, Newell went on, computer system levels are a reflection of the nature of the physical world. "They are not just a point of

The Hagia Sophia I referred to in Chapter 2 is another instance of the search for general laws of intelligence.

This last great intellectual effort, Soar, had such ambitious goals that Newell's premature death kept him from seeing it develop fully. (In his last illness, he wondered to Joe and me if his fatal cancer had arisen from the days of his naval service, where, from a very few miles distance, he witnessed the atomic bomb tests on Eniwetok.) Although researchers continue to work on Soar and models like it, grand models of the complete suite of human cognitive behaviors are not yet at hand. For one thing, they demand the utmost of human intelligence to fill them out and get the details right.

As it happens, about the time Newell was giving his William James Lectures, a brilliant researcher at the University of Toronto, Geoffrey Hinton, was exploring a part of AI that had lain dormant, was even presumed dead when Marvin Minsky and Seymour Papert had seemed to say all there was to say about it. This was neural networks. Minsky's and Papert's model was vastly a simplified version of the brain with only an input layer and an output layer. In 1986, Hinton showed that a technique called backpropagation could train a deep neural net, one with more than two or three layers. Much more computing power was needed before Hinton and two of his colleagues could show that deep neural nets, using backpropagation, dramatically improved upon old techniques in image recognition. This has led to deep learning, and applications that propagate like mayflies, including nearly unerring human facial recognition by computer, talking digital assistants, and not incidentally, the 2018

view that exists solely in the eyes of the beholder. This reality comes from computer system levels being genuine specializations rather than being just abstractions that can be applied uniformly" (Newell, 1990).

Turing Award for Hinton (vice president and engineering fellow at Google, chief scientific advisor at the Vector Institute, and professor at the University of Toronto), Yann LeCunn (vice president and chief AI scientist at Facebook and a professor at NYU), and Yoshua Bengio (professor at the University of Montreal, and science director of Quebec's AI Institute and the Institute for Data Valorization).

The time that a computer exhibits a grand suite of complete human cognition behaviors may be approaching. In January 2016 MIT and Harvard sponsored a day-long symposium called "The Science and Engineering of Intelligence: A Bridge across Vassar Street." Vassar Street separates MIT's computer science and AI research from the Broad Institute and other Cambridge neuroscience research centers. The symposium's aim was to show how AI and neuroscience have critically influenced and inspired each other and how quickly we're learning about each.[7]

Skeptics exist. Ed Feigenbaum, for one, believes a grand theory of mind can't happen for years, maybe never, because human intelligence grew in such a contingent, biologically opportunistic way. He strongly believes that intelligence in machines will come from the bottom up, not the top down, and incrementally.

Neuroscientists are more sanguine. But their task is mighty, and they could be wrong. Stuart Russell, a professor of computer science at the University of California Berkeley and coauthor of the textbook *Artificial Intelligence: A Modern Approach*, recently said that, although we know how to make the computer do many things humans can do, we haven't yet put them all together in a working grand

[7] Remember from Chapter 1 that Demis Hassabis of DeepMind argued that his company's program, AlphaGo, is the opening to general as opposed to specialized artificial intelligence, which implies a unified theory of cognition.

scheme—and maybe, he added, that's a good thing. He seemed to imply that this might court a dismal fate. In 2017, as if to make the point another way, he gave a lecture where he presented a working example of a small, cheap killer drone that "could wipe out the population of half a city," a drone "impossible to defend from." *Impossible* is a big word. Should the quest for general AI therefore be abandoned? I don't think Allen Newell would agree. Newell strongly believed that grand working schemes—an overarching question that drives a personal scientific agenda, which in his case was understanding the human mind—is exactly how science should be done.

The MIT Group

AI by Hook or by Crook

1.

In March 1975, I began a series of interviews for my proposed history of AI with pioneers around Cambridge, Massachusetts. Among the first I interviewed was Marvin Minsky, one of AI's four founding fathers, along with John McCarthy, Allen Newell, and Herbert Simon. Minsky was welcoming and deeply generous with his time. By then, Minsky had already won the Turing Award and would go on to win many more honors, including the Japan Prize in 1990, the International Joint Conferences on Artificial Intelligence's Award for Research Excellence in 1991, and the Benjamin Franklin Medal from The Franklin Institute in 2001. He even consulted on Stanley Kubrick's groundbreaking movie, *2001: A Space Odyssey*, with explicit credits.

Everybody agreed that Minsky was one of the smartest people on the planet, but what few mentioned was his appealing generosity of spirit. This might be why the list of his students is an impressive roster of scientists who've made their own dazzling contributions to AI and other computing areas. "I don't think of myself as a teacher," Minsky once said to me. "I'm more like a gardener. I let the plants grow, I nourish them, and I weed the garden." By that he meant that he encouraged creativity and gently (or maybe not so gently) guided his students along paths that would allow their creativity to flower.

Another example of Minsky's generosity of spirit. We were chatting about an early worker in AI who'd had one great success and then failed to do more. "Ah," he said quietly, "we don't know what circumstances in people's lives might bottle them up. It isn't necessarily failure of intelligence. Things just happen." It was a reminder not to judge so quickly.

2.

Marvin Minsky was born in New York City and attended Ethical Culture Fieldston School and The Bronx High School of Science. He came from an established New York family, and when, in the mid-1980s, he told his elderly mother that he was about to publish *The Society of Mind* with Simon and Schuster, she murmured thoughtfully, "Liked Simon. Never liked Schuster."

I reported in *Machines Who Think* how the young Minsky, as an undergraduate at Harvard, fed his curiosity by going to all the teas before or after talks on topics of every description. (He fed his youthful appetite too, scarfing up the cookies.) He understood what most shy undergraduates do not: generally people are delighted to talk about their research with anyone, even an undergraduate, who shows a polite interest. Although he was nominally doing mathematics (he overlapped at Princeton in graduate mathematical studies with McCarthy and Newell, but none of them knew each other well then), he was interested above all in the questions surrounding intelligence. He'd been deeply influenced by Warren McCulloch at MIT, who did early studies on neurons, and Minsky's

PhD dissertation was a mathematical model of certain neural functions in the brain. He visited Bell Labs in the summer of 1955, where, with Claude Shannon's blessing—Shannon the father of information theory—he and McCarthy dreamed up the whole idea of a conference the following summer at Dartmouth of people who suspected these new machines called computers could be made to think.

After the Dartmouth conference, Minsky was still formulating how AI might be achieved and wrote the first of many versions of what would come to be called "Steps Toward Artificial Intelligence." He admired what Newell and Simon had done, but thought he was no longer interested because Newell and Simon were constructing models of *human* intelligence. Instead, he wanted to achieve machine intelligence in any way possible.

With Seymour Papert, Minsky wrote an influential, if difficult book, *Perceptrons* ("We didn't leave enough easy problems for graduate students to tackle," he laughed—although much later when computing power was up to the task, the book would be seen as a forebear to deep learning). He also continued to cultivate his graduate students, whose achievements were signal, and invented new theoretical approaches to achieving machine intelligence. Besides *Perceptrons*, in 1969, which was subsequently expanded twice, he wrote a book on frames, a computational structure for laying out facts about objects and events—in other words, knowledge representation—that significantly influenced AI program design.

Meanwhile, Minsky invented some important instruments, such as a precursor to the laser-scanning microscope, and an early graphical display. One of his most famous instruments was the Logo turtle, developed with Seymour Papert. This was a robot that executed the instructions of children who were learning the simple but powerful programming language Logo, Papert's creation. In the mid-1970s, I spent several hours watching Boston–area eight- and nine-year-olds at computer keyboards, their faces radiant with mastery, as they instructed the turtle to move around the floor.

But gradually, Minsky came back to examining human intelligence because that had been his original impetus. Even in the 1970s, he laughed with me that it had only *seemed* as if Carnegie Mellon and MIT had gone different ways. In fact, they were both interested in understanding and modeling human intelligence as the best proof of concept, as engineers like to say. His later books, *The Society of Mind* and *The Emotion Machine*, testify to exactly that. They also testify to Marvin Minsky's significant contributions as a theoretician of intelligence, human or machine.

Yet a theory wasn't enough, computer scientists like Newell and Simon would grumble. You needed experimental evidence to prove or disprove it, to refine or expand it. This was a friendly but persistent difference between the two schools of thought.

3.

Although Minksy and I taped numerous interviews at MIT,[1] in my journal I mentioned a couple of visits to his Brookline home.

February 7, 1977:

> *Spent most of the day with Marvin at his house, and if I hadn't been ready to turn into a block of ice by the time it was all done, I'd have been better*

[1] These and all other interviews I conducted for Machines Who Think are available in the archives of Carnegie Mellon University.

company. The house deserves description. A large house, many rooms, each one lined, stacked, stuffed with memorabilia, objets, such as: a large harmonium (which looks like an organ to me), a jukebox, dolls, piñatas, odd chairs and sofas. In the family room is an impressive amount of sound equipment, a piano, games, records, a human arm attached to the wall with significant bones painted red, white and blue, a trapeze suspended from the ceiling beams, various mirrors, including a searchlight mirror, two mirrors from a telescope, and a couple of concave mirrors fit on top of one another so they look like a large wok. Their function is to reflect a little metal frog, who sits on the lowest mirror, up into a hole in the upper one, thus giving you the impression you have a solid metal frog suspended in the interior, which you can put your finger through quite easily if you've a mind to.

Here, after we'd talked AI for a while, Marvin gave me a sample of his new love, which is composing music. Now of all the kinds of music there are, I wouldn't have expected Marvin Minsky to compose this, but out it comes, one beautiful, fluid Bach-like fugue after another. I was enchanted. And told him so. The melodies were lovely, lyrical, beautifully realized and then counterpointed. If he'd told me old J.S. himself composed them, I'd have believed it. Then he played some Prokofiev-like music, and finally some music for children, all of it, it seemed to me, exceptionally fine. I was surprised that I liked it so much at once, but it had a natural grace that spoke to me directly. We talked about composing and he told me he simply put down the music he heard in his head—the relationships weren't (necessarily) mathematical but were discovered after the fact.

A brief lunch, and we spoke some more. Then Gloria Rudisch, Marvin's wife, who is both a pediatrician and the health officer for the City of Brookline, came home "with a robot for Marvin to fix." She's a small, stocky woman, black hair in a neat pageboy, and she was almost overwhelmed by the suitcase she was toting. When she opened it, a hand and sneakered foot fell out. She extracted a very lifelike woman, dressed in a blue jogging suit, rigged up in such a way you could measure on a meter whether you'd "restarted her heart," or "restarted her

> *breathing," by mouth-to-mouth. I'm hard put to describe the picture of Marvin and Gloria working furiously over this mannequin to get her prepared for a class Gloria was about to teach, this life-sized and so lifelike stiff lying on the couch, the dog and me riveted by the whole affair. Gloria skittered out at last with the suitcase, and Marvin carried the dummy under his arm to the car. Not a sight I'll soon forget.*
>
> *Marvin is very serious about his composing, wonders if he should just make the big break and change his life altogether. If I hadn't been so cold, I could've gone on for a long time. I don't know whether the Minskys keep the heat down because they're good citizens, think it's good for our health, or they're just indifferent. An hour in the semi-tropical heat of "the fine old fellows"* [Joe and I were staying at Boston's Harvard Club, and the house manager used that phrase to describe our elderly fellow guests, to explain why the heat was so high] *and I'm still not thawed out, but grateful indeed for the old fellows' terrible circulation which keeps everything nearly molten.*

A few days later, the Minskys invited Joe and me to dinner with a large, congenial group. February 10, 1977:

> *Dinner tonight chez Minsky, cooked by Gloria and also by Seymour Papert. I had a long talk with Seymour's friend, Sherry Turkle,* [later, for a while, his wife, and to become a celebrated investigator of human behavior with computers] *who's doing a sociological study of why computer scientists do what they do, having just completed a study of French psychoanalysts, called* French Freud.[2] *Also at table were Felix, Marvin's friend since grammar school, Albert Mayer, an MIT professor acting as our social secretary for the week, and Marvin's son, Henry, perhaps fourteen, who complained to me about having to read Jane Austen in school when he'd rather be reading Kurt*

[2] What a great title! When I rediscovered this in my journal, I asked Sherry Turkle why she'd changed it to the bland Psychoanalytic Politics. "The publisher," she replied. "They thought it might be misunderstood, or confused with another book. I was very young and didn't know better." Weren't we all, I agreed.

Vonnegut. "They're doing the same thing," I said, "social satire." I'm not sure he was convinced.

4.

For Minsky and his students, robotics raised fundamental issues. How did a dumb video camera, connected to a dumb contraption that served as an arm, connected to a computer, produce intelligent behavior? How did the arm understand that it was being asked to pick up building blocks and move them from one place to another? This stood for one of the central questions about intelligence: How does intelligent behavior emerge from dumb tissue, or dumb components of any kind? (It was nearly half a century later before we began to get answers to those questions—an elaborate set of reciprocal signals between brain and limb.)

In the early 1970s, Minsky and Papert began formulating what would become Minsky's 1988 book, *The Society of Mind*, at the time a somewhat speculative, but persuasive, and finally influential, set of theories proposing that all minds, natural and artificial, were made up of small unintelligent components. Yet acting in concert, sometimes using well-tested algorithms, sometimes using rules of thumb, they produced what we call intelligence. This is a common assumption now, an early exploration of the phenomenon of emergence, but the book caused a tremendous stir among brain scientists, psychologists, and philosophers, who were laboring toward something more elegant in the way of a grand unified theory of human intelligence.

Almost twenty years after *The Society of Mind*, Minsky turned to what we call emotions. Could he account for the role that emotions play in intelligence? Given the distinction Western culture has always made between reason and passion, did emotions play any role at all in intelligence? He came to believe that this distinction, asserted since the classical Greeks, was simply wrong.

In a 2006 book called *The Emotion Machine*, Minsky proposed that emotion plays a vital role in intelligence. In *The Society of Mind*, he'd argued that agents in the mind worked together toward goals. Now he changed the concept of agents to resources, because the word *agent* misled readers into thinking that a person-like thing—a homunculus, so to speak—existed in the brain and could operate independently or cooperate with other agents, in much the same ways people do in the real world. On the contrary, he said, most resources in the brain are specialized to certain kinds of jobs and cannot directly communicate with most of the brain's other resources.

In *The Emotion Machine*, he argues that our longtime distinction between passion and reason rests on misunderstanding both terms. Passion and reason each are probably a hundred different things at least, the consequences of the behavior of tens of thousands of inherited genes, their expressions raw and uncontrolled, until we mature and learn to control them. Many of these resources are inaccessible to deliberate scrutiny, for we've overlaid other processes on them as we've matured.

For convenience, or from laziness, we use what Minsky calls "suitcase words," like *love*, *hunger*, *anger*, *suffering*, and *pleasure*, as if they had precise meanings. Instead, he argues, each suitcase word has many different items stuffed into it as we attempt to describe large networks of processes inside our brains. Consciousness, for example, refers to more than twenty such processes. "Each of our major 'emotional states' results from turning certain resources on while turning certain others off—and thus changing some ways that our brains behave" (Minsky, 2006.).

As a rule, emotions are ways to think that increase our resourcefulness. This is vital. If a program worked only one way, it would get stuck when that one method failed. "The resourcefulness of the human mind comes from having multiple ways to deal with things—no matter that, from time to time, this causes bad things to happen to us" (Minsky, 2006). Even our sense of self is impermanent: we have multiple models of the self and switch between them as we learn when it's useful to do so (Simon, 1991).

Although the *The Emotion Machine* presents a different way of thinking about the role of emotion in intelligence,[3] it grows out

[3] An answer to that question was only to come more than half a century later, in a collaboration between Caltech neuroscientists and roboticists. They devised a robotic arm, a prosthesis, equipped with a brain-machine interface that can read and respond to the intentions of its human patient, a man otherwise unable to move his arm owing to an old gunshot wound. The scientific team showed that an elaborate set of messages travels from the brain (in this patient's case, implanted with sensitive electrodes) to the appendage, and back in a rich feedback system. Richard Andersen, "The Intention Machine." Scientific American, April 2019. A similar system is under construction jointly between the University of California, San Francisco and the University of California, Berkeley, for brain messages to cause speech. Carey, Benedict. "Scientists Create Speech from Brain Signals." The New York Times, April 24, 2019. Much psychological literature, especially popular reading, had treated emotion as distinct from intelligence, sometimes a separate kind of intelligence in its own right. That view has been hotly contested and is different from Minsky's more integrated role for emotions in intelligence. The March 2014 issue of Global Advances in Health and Medicine includes a long paper, "Emotion: The Self-regulatory Sense," by K. Peil, who says that emotion, broadly construed, plays a fundamental self-regulatory role in any organism. In the April 2015 issue of Scientific American, the article "Conquer Yourself, Conquer the World" by Roy F. Baumeister discusses the complicated role self-control plays in human behavior. For a focus on hatred specifically, see "The Point of Hate" by Anna Fels in The New York Times, April 14, 2017. Brain scientists generally agree that emotions play a key role in individual decision-making, but the current model suggests that networks in the brain compete for supremacy, with emotions often winning over reasoning, because emotions are a fast, economical way of deciding and help lift the daily cognitive load.

of what Minsky had learned over a lifetime's research in AI. Both *The Society of Mind* and *The Emotion Machine* are lucid expositions of ideas that are current in—or at least not alien to—both brain and AI research. He also called on findings from psychology, animal behavior, cognitive science, and genetics (a substantial part of our behavior is endowed in our genes).

Several AI researchers conceded that Minsky might be right, but where were the computer programs that instantiated these ideas, that separated science from mere conjecture?

A partial answer comes from Minsky's MIT colleague, Rosalind Picard, who had already coined the term *affective computing.* She too argued that reasoning and emotion were inseparable, and emotions were necessary for true machine intelligence. Picard, along with her graduate student, Rana el Kaliouby, began testing software that could read emotions on the human face. They formed a company called Affectiva to sell the systems, but their customers, instead of being clinical researchers in autism, say, were overwhelmingly market resesarchers who wanted to use the software to refine products and advertisements. Picard stepped away from this as too distant from her original medical goals, but Kaliouby stayed with Affectiva, now a thriving business of reading human emotions for its international clients. To the train the software, Affectiva began with a handful of actors and now has massive amounts of data. This has refined the program's skills to the point where it's more sensitive at reading emotions than most humans.

Meanwhile, Picard has pursued the brain-mind-body connection along multiple fronts. One helpful wrist device she helped develop reads brain and body electrical signals, allowing epileptics to anticipate a seizure twenty minutes before it takes place. “We want to give individuals something to help them do better, rather than just focusing on AI that only people in powerful positions have access to.” She now studies healthy people, to see how they maintain their wellbeing. “In the world of AI, some of us are stepping back and asking what are we doing to human health. What leads to true human flourishing and wellbeing? Are we enabling the kind of AI that gives wealth and power to a smaller and smaller number of people? Or are we enabling AI that helps people?” (Wapner, 2019).

Many people, Minsky writes, have come to accept that the human brain is an electrochemical organ, but they still believe that a mystery will always remain about how a living thing could ever result from nothing more than material stuff, whether synapses or electrons. “That once was a popular belief, but today it is widely recognized that behavior of a complex machine depends only on how its parts interact, but not on the ‘stuff’ of which they are made (except for matters of speed and strength). In other words, all that matters is the manner in which each part reacts to the other parts to which it is connected” (Minsky, 2006).

In machine or human brains, these resources are proving to be hierarchical networks of processes (again, the central idea of Allen Newell’s Soar model), many of the lowest systems not even available to the higher systems. Your conscious mind can’t access the processes that keep you steadily breathing or standing upright, for example, though they’re basic to your existence. In humans, mapping just where these processes reside in the brain is one of the great goals of present-day brain science.

Minsky (2006) observes: “Exploring, explaining, and learning must be among a child's most obstinate drives—and never again in those children's lives will anything push them to work so hard."

Minsky proposed a group of hypotheses still to be fully validated. Neuroscientists had already begun such exploration as he wrote *The Emotion Machine,* and they continue. Even now, no one knows whether Minsky's ideas are correct in general or in particular. We do know that emotions are finely nuanced and contain a wide variety of fleeting, sometimes contradictory aspects. Machines can read human emotions and respond to them, whether they're evaluating audience responses to TV pilots, guiding autistic individuals through a world of affect that puzzles them, or assisting a digital nurse to evaluate a patient (Stone & Lavine, 2014).[4]

That emotions are a fundamental resource already integrated into intelligence, not merely to be ignored, suppressed, or overcome, has an appealing economy. Individual maturation involves learning how to control these potent fundamental resources. Oxford philosopher and cognitive scientist Nick Bostrom (2016) argues that such maturation must take place with AIs too, and perhaps this is so.

5.

In the fall of 2013, I was lucky to sit in on the first weekly meetings of the Center for Brains, Minds, and Machines at MIT and Harvard,

[4] A special issue of Science called The Social Life of Robots has many articles that cover robots as coworkers, neuromorphic robots, the challenge of robot sensors, giving robots the big picture of the world, the psychological implications of robots that look human, robots and the law, and robots in biological research. Yes, I find emotion-reading robots creepy. But that's a personal reaction, which may or may not be germane to AI's future research. If there's one thing I've learned, it's that a thinking-fast reaction needs much more thinking slow to properly examine it. Suppose, for example, an emotion-reading robot becomes a pedagogical tool that teaches humans how to understand and respond better to the emotions of people around them.

meetings that continue. I listened to scientists in each of those fields offer to one another a brief description of their work. One afternoon began with what we know about how humans understand scenes. Another scientist described how humans recognize scenes (slightly different from understanding a scene). A third scientist presented findings of experiments with a brain imaging technique, where the scientist shows her subjects an image and then decodes the brain waves. A fourth scientist offered a means of teaching machines common sense via storytelling.

At the end of their presentations, each scientist added: if my models, questions, or answers are useful to you, use them. If you think I can help you, get in touch. Get in touch anyway.

During these openhanded afternoons, scientists across disciplines tried to help each other understand what intelligence is. Right now, this kind of exchange is taking place all over the country and the world. The challenge is enormous, and the investigative instruments are barely up to it, though they'll surely continue to improve. Minsky made no apologies from the outset. Years ago he said to me, "Look how long physicists have been studying physics. Do we think the brain and mind are less complicated?" E. O. Wilson says decisively: "The human brain is the most complex system known in the Universe, either organic or inorganic" (Wilson, 2014).

The brain's energy efficiency is one complexity that scientists have yet to understand. David Cox, a professor of molecular and cellular biology and computer science at Harvard's Center for Brain Science, points out that the human brain has the capacity for tens of petaflops yet consumes only 20 watts of power. (A petaflop is a measure of supercomputing speed; one peta equals a quadrillion floating point operations per second, or *flops.*) Current supercomputers have arrived at the tens of petaflops, but their appetite for power is gargantuan—just getting rid of the heat they generate is a challenge.

The brain can solve problems we don't know how to program computers to solve, regardless of the power those computers can muster. That doesn't mean we won't ever know. But we don't know now. I asked Tomaso Poggio, the head of MIT's Center for Brains, Minds, and Machines, which set of researchers, neuroscientists, cognitive psychologists, or computer scientists, was likely to develop—or discover—the mechanisms of intelligence first. "It's a race," he replied, smiling.

6.

On one of my early visits to Cambridge in the 1970s, I interviewed Ray Solomonoff, one of the original attendees at the Dartmouth conference. Ray's fan-like beard was already gray, and he'd lost much of the hair atop his head. Behind his glasses, his eloquent eyes seemed spiritual in their intensity. He was very much a free spirit, still doing mathematical modeling of mind, but attached to no institution. After we talked, he and his girlfriend offered to take me out to forage for salad greens in Harvard Yard.

After the Dartmouth Conference, Solomonoff's work fell into eclipse for several decades, but in the mid-2000s, was revived in a subfield called artificial general intelligence, where researchers sought a universal way of learning and acting in any environment. This pattern of eclipse and revival has happened several times in AI (recall Newell and Simon's General Problem Solver) where original good ideas, impossible to implement with the technology of the time, suddenly become possible, and even better, useful. Deep learning is a grand example.[5]

Oliver Selfridge, officially at MIT Lincoln Laboratory (also known as Lincoln Labs) but with a post as associate director of MIT's Project MAC[6] in the early 1960s, was another early advocate of an integrated approach to AI. He'd been working on pattern recognition and machine learning—a presentation he made had electrified Allen Newell at RAND in the mid-1950s—and Selfridge's 1959 paper, "Pandemonium," a proposal for machine learning, is considered a classic in the AI literature. Selfridge coined the term *intelligent agents* for autonomous software capable of sensing and responding to changes in their environments, an idea that would develop more fully in later years (Feigenbaum & Feldman, 1963). In the mid-1970s he was also seeking an approach to general intelligence and was disappointed, he said, that pattern recognition had been pushed off to

[5] Computer science on the whole is regrettably ahistorical. An eager researcher will gladly reinvent the wheel before he'll take the time to search the literature and see if anyone else has tried what he has in mind. Acknowledging this, Manuela Veloso, an eminent roboticist at Carnegie Mellon then, exploded to me, "Such a waste!" But William A. Wulf, for eleven years the president of the National Academy of Engineering and a computer scientist himself, says this allergy to history reflects the way funding is appropriated and papers are selected for publication; only the new matters, whether or not it's actually new. Unlike mathematics, with its longstanding cultural traditions to cite precedents, computer science in general has no such pressure. Raj Reddy, maybe to tease me, said dismissively, "Oh, it's just easier to reinvent than try and track down some original idea." To finger these reinventions requires a canny practitioner-turned-historian with breadth and depth, like Nils Nilsson, in his The Quest for Artificial Intelligence (Cambridge University Press, 2010). However, Professor Mary Shaw has informed me that at Carnegie Mellon, the introductory course for new PhD students in software engineering begins with about two dozen classic papers that every software engineer should know, and each unit of the course bridges from some of the fundamental papers to how the ideas have evolved. Those early papers account for about a third of the course reading. "We introduced this in a curriculum revision a few years ago because we were frustrated about exactly this problem." (Private communication)

[6] The acronym stood for a number of phrases, including Mathematics and Computation, Man and Computers, and so on.

be its own subfield, unrelated to mainstream AI. This too was to be slowly reversed, but only after some decades.

For *Machines Who Think*, I also visited the elusive Claude Shannon, best known for his work on information theory, the theoretical foundation of the digital revolution.[7] He'd allowed the Dartmouth conference to take place under his aegis with the understanding that John McCarthy and Marvin Minsky would do the work. In his seventies, Shannon was a prepossessing man, his features finely modeled, courtly and soft-spoken, happy to talk about early times at both Bell Labs and MIT. He'd retired from MIT, so no longer rode his unicycle around the academic halls, but he was still full of playful and intellectual verve.

Shannon then lived in a grand old Victorian house in Somerville, with sweeping views of the Boston skyline. After our interview, he took me into another room to see the remains of a legendary maze that a mechanical mouse called Theseus had run through in 1950, part of a very early experiment in machine learning. Years after I interviewed Shannon, Joe was stunned to see him as a new inductee into the National Academy of Engineering. Shannon should have been a member for decades, having already won the National Medal of Science, among many other honors. Sadly, he eventually suffered from Alzheimer's and died in a Massachusetts nursing home in 2001, oblivious, his widow said, to the wonders he'd helped bring about.

[7] Shannon would tell Joe and me at a 1984 conference in Brighton, England, that he'd tried to get people to call what he did communications theory and not information theory, but the name stuck. "Let's start a campaign to rename it," Shannon joked to us, knowing how impossible it now was. Joe soon found an early paper of Shannon's where he'd used the term information theory himself, setting the precedent.

Edward Feigenbaum

A Bashert Friendship

1.

Edward Feigenbaum, a prominent member of the second generation of AI researchers and an academic son of Herb Simon, took AI research in the opposite direction from his forefathers. This was the Tristan chord I'd been deaf to when I worked for him. Because I didn't know what had come before, I couldn't know how radical his departure was.

The second generation of AI researchers departed from their forefathers by being less interested in modeling precisely how human intelligence works than in devising ways to help humans accomplish things—as you'll see with Feigenbaum and, in the next chapter, Raj Reddy.

Feigenbaum was born in Weehawken, New Jersey, on January 20, 1936, in the heart of the Great Depression. While he was still a young boy, his father died, and his mother remarried to Fred Rachman, an accountant for a baked goods firm. The boy and his stepfather developed a warm relationship, and his stepfather would take him faithfully each month across the Hudson to New York City to see the show at the Hayden Planetarium. ("They did new shows once a month in those days," Feigenbaum recalls.) Then they'd add a visit to one or more rooms of the American Museum of Natural History. These visits got him started as a scientist.

Fred Rachman often brought work home, and a mechanical (soon, an electromechanical) calculator to do it. The boy loved these Marchants and Fridens, and learned to work them skillfully. "I didn't have a letter on my sweater, but I could lug these calculators on to the bus to school, and show all my friends what I could do with them."[1]

From Weehawken High, Feigenbaum went on a scholarship to Carnegie Institute of Technology (now Carnegie Mellon) to study electrical engineering. Money was tight: he often had to work outside school to help support himself. One of those jobs was teaching science in a Lubavitcher elementary school in Pittsburgh's Squirrel Hill. "I couldn't mention sex, I couldn't mention evolution, I couldn't mention a whole bunch of things that the rabbi forbade," he laughed once. "Teaching science under those circumstances was a challenge."

As a sophomore in electrical engineering, Feigenbaum felt "something was missing." He found a graduate-level course called Ideas and Social Change, taught by the behavioral scientist James March. March allowed Feigenbaum into the course, where he learned about John von Neumann and Oskar Morgenstern's *Theory of Games and Economic Behavior*. Feigenbaum loved it. Soon, modeling of behavior was introduced, even more fascinating to the

[1] Feigenbaum would return the favor of those planetarium visits and calculator loans. Years later, in the mid-1960s, when Fred's job looked precarious because industry was shrinking in New York City, Ed brought his stepfather to Stanford to learn how to be a computer operator, switching the tapes on tape drives and watching the console lights that signaled to operators the steps they needed to take.

undergraduate. That summer, March gave Feigenbaum a job doing experiments in social psychology, which led to his first published paper with March, on decision-making in small groups. March also introduced Feigenbaum to the senior colleague with whom he was writing a book on organizations, Herbert Simon. Simon took an interest in the youngster and helped him get a summer student fellowship the following year. Feigenbaum subsequently enrolled in Simon's course called Mathematical Models in the Social Sciences. This was the course where Simon announced, "Over the Christmas holidays, Al Newell and I invented a thinking machine."

Feigenbaum would later call that a born-again experience. He took the IBM 701 manual home and, by dawn, was hooked on computers. In graduate school, his PhD dissertation, written under Simon's supervision, was a computational model of some aspects of human memory, Simon's great preoccupation. "Here's the data," Simon had said, showing him what the psychology literature had carefully accumulated by experiments. "Let's make sense of it."

Feigenbaum remembered later, "Never, ever was the *brain* brought up. This was altogether a model of the *mind*, of human information processing with *symbols* at the lowest levels." (McCorduck, 1979)

Psychologists had collected much data on how people memorized lists of nonsense syllables. Could Feigenbaum write a computer program that remembered and forgot the same way that people did, and thus explain the behavior? He could. In memorizing lists of nonsense syllables, he realized, people didn't memorize whole syllables. Instead, they memorized tokens that stood for the syllable, tokens that then called up the entire memory. He incorporated this and other memorizing and forgetting patterns in a groundbreaking program called Epam, for Elementary Perceiver and Memorizer, but also because at the time Simon was studying the Theban general and statesman, Epaminondas. Simon would eventually take this work further with psychology colleagues, but by then, Feigenbaum was in pursuit of something more interesting.

Between Feigenbaum's PhD and his first academic post, he took a year to visit the National Physical Laboratory in Teddington, England, and then came to Berkeley, where he and his friend from Carnegie Institute of Technology, Julian Feldman, taught organization theory and artificial intelligence and where Feigenbaum and I first met. When he and Feldman saw how eager students were to know more about the topic of AI and its growing importance, they knew a textbook was needed, and thus was born *Computers and Thought*, the first collection of readings in the field.

And so was our friendship. To write of friendship is to consider the sweep of a lifetime's respect and affection. Such a friendship, Montaigne observes, has no model but itself and can only be compared to itself. In 1960, Ed Feigenbaum had detected in a young Berkeley co-ed something out of the ordinary (or so it felt to me, that young co-ed). He and Julian Feldman invited me to work on *Computers and Thought*, my introduction to the field. When I left the field for other interests, I often returned to Ed to hear what was new in AI. But the friendship endured, with great depths that transcended anything professional. For that, I've always been grateful.

Years later, I'd reflect on how much Ed Feigenbaum is a man who loves women. He has two beloved daughters from his first marriage. His second marriage to Penny Nii, a Japanese-born woman who became his scientific colleague, brought him two beloved stepdaughters. He's been drawn to strong, imaginative women, and made sure the women around him, in his family, in his research groups, flourished magnificently. All of those women went on to singularly successful careers. To me, he was teacher, mentor, big brother, and finally, beloved friend.

So Feigenbaum and I got along smoothly and happily with each other from the outset. Once during the *Computers and Thought* days, Feldman walked into a small office where Feigenbaum and I were chatting, listened to us for a moment, and shook his head. Oy, such *yentas*!

I shrugged. Yes, Feigenbaum and I loved to talk to each other about everything under the sun. Feldman nodded. It was, he said, *bashert*. That sent me to a Yiddish dictionary: *foreordained, fated*. So it was.

2.

After *Computers and Thought* was delivered to the publisher, I moved on. Five years later, when Feigenbaum went from Berkeley to Stanford, he called me to come and join him as his assistant, which would change my life.

I learned. I watched. I absorbed. I asked questions—always patiently and fully answered. I didn't know that, at this moment, his hands plenty full with running the Stanford Computation Center, not to mention the serious sailing he was doing on San Francisco Bay and beyond the Golden Gate, he yearned for something more ambitious for AI. It was coming to him that nothing was bigger than induction.

"Induction is what we're doing almost every moment, almost all the time," Feigenbaum said. We continually make guesses and form hypotheses about events. Brain scientists believe that at the level we do it, this is a uniquely human speciality, but in the 1960s, Feigenbaum was only asking how induction works in scientific thinking. Here was a significant challenge for AI, more ambitious, certainly more important, than how people memorized lists of nonsense syllables. Was the field ready to tackle something so sophisticated? Was he?

By chance, Feigenbaum encountered Joshua Lederberg, a Nobel laureate in genetics at Stanford, and told the geneticist what kind of problem he was seeking. “I have just the thing for you,” Lederberg said. “We’re doing it in our lab.” It was the interpretation of mass spectra of amino acids, the task of highly trained experts. Lederberg was heading a project for a Mars probe to determine whether life existed on Mars but knew he couldn’t ship human experts to operate mass spectrometers on the Red Planet.

In 1965, Feigenbaum and Lederberg gathered a superb team, including philosopher Bruce Buchanan and later Carl Djerassi (one of the “fathers” of the contraceptive pill) plus some brilliant graduate students who would go on to make their own marks in AI. The team began to investigate how scientists interpreted the output of mass spectrometers. To identify a chemical compound, how did an organic chemist decide which, out of several possible paths to choose, would be likelier than others? The key, they realized, is knowledge—what the organic chemist already knows about chemistry. Their research would produce the Dendral program (for dendritic algorithm, tree-like, exhibiting spreading roots and branches) with fundamental assumptions and techniques that would completely change the direction of AI research.

As Richard Wagner’s celebrated Tristan chord changed all subsequent musical composition, Dendral changed all subsequent AI. Until Dendral, the most important feature of AI programs was their capacity to reason. Yes, the earliest programs knew some things (the rules of chess, the allowable rules of logic) but emphasis had always been on reasoning: refining and elaborating the way the program moved toward its goal. Hadn't the great Aristotle called humans the reasoning animal? Wasn't this confirmed by nearly every philosopher who ever thought about thinking? This unquestioned assumption led Allen Newell and Herb Simon to design the General Problem Solver program, which tried (but mostly failed) to solve problems generally.

More than two thousand years of philosophy was wrong. *Knowledge*, not so much *reasoning*, was essential. You can almost hear the protests from the shades in the agora.

Although Dendral's reasoning power, what would come to be called its inference engine, was strong, Dendral's real power and success came from its detailed knowledge of organic chemistry. Knowledge allowed the program to plan, put constraints on possible hypotheses, and test them. As a stand-alone program, Dendral became essential to working organic chemists. Its heuristics were based on judgment and specific chemical knowledge, what in humans we call experience and intuition. Joel Moses at MIT would say to me later, "It's insane to think you can do brain surgery without knowing anything about the brain—just reason your way through it."

The knowledge principle, as Feigenbaum came to call it, asserts that specific knowledge is the major source of machine *and* human intelligence. With the right knowledge, even a simple inference method will suffice. Knowledge can be refined, edited, and generalized to solve new problems, while the code to interpret and use the knowledge—the reasoning, the inference engine—remains the same. This is one reason why, in the last few years, AI has become noticeably smarter. The amount of knowledge on the Internet available to Watson, Google Brain, or language-understanding programs (or scores of startups) has grown dramatically. Big data and better algorithms implemented at multiple processing levels have vastly improved performances. But even the field of machine learning, dependent as it is on algorithms, acknowledges that domain knowledge is essential to intelligent behavior.

Dendral, Buchanan said, was the first program to attempt to automate scientific inference. It was the first program to rely on textbook knowledge *and* the knowledge of human experts in a scientific domain. Dendral was the first program to represent such knowledge in an explicit and modular fashion. "We were learning how to represent the knowledge in a nice, clear, high-level symbolic way—you could actually see what the knowledge was," Feigenbaum added. (This idea was to be significant in the future digital humanities.)

It didn't matter that it was knowledge already known: patterns of remembering and forgetting nonsense syllables had also been well documented when Feigenbaum sat down to write Epam. Those empirical experiments verified that the program successfully imitated human learning and forgetting in one small domain. Now he and his colleagues had set out to model the process of spectra interpretation well enough that a computer program would match or exceed what a human expert could do. But Buchanan, the trained philosopher on the team, whose interests were in scientific discovery and hypothesis formation, was eager for Dendral to go further and make discoveries on its own, not just help humans make them. In decades to come, this would happen, but by then, other scientists had taken up the challenge of a progam that makes scientific discoveries on its own, as we'll see later.

The team discovered that the more you trained the system, the better it got. Because it embodied the expertise of human specialists, this kind of program came to be known as an *expert system*. In short order, Dendral was followed by Mycin, a program to help a physician identify and recommend antibiotics for infectious diseases. If asked, Mycin could also explain its line of reasoning. Another later program, Molgen, generated and interpreted molecular structures.

If Dendral and later Mycin came to outperform human experts, Molgen had a different challenge. Not much was known about generating and interpreting molecular structures, and that modest knowledge was stored in the heads of human experts around the world. To store and draw on that geographically distributed knowledge, the Molgen program ran on the only non-ARPA-funded machine allowed on the ARPAnet, the precursor to today's Internet. Users could dial in from all over the country—university biology departments, pharmaceutical companies—to access the Stanford sequence manipulation routines and add their own knowledge. Before long, some 300 users were coming in over the ARPAnet.[2] But it was another twenty years until computer graphics and networks were up to the task of generating wide-scale automatic molecular structures. Molgen thrives worldwide now.

The first major step in constructing an expert system was to interview human experts and gather their specialized knowledge. Next, that

[2] Molecular biologist Larry Hunter has argued persuasively that molecular biology simply cannot be done without AI techniques that verify knowledge, trace lines of reasoning, and keep ontologies (agreed-upon knowledge) straight and consistent.

knowledge had to be cast in executable computer code. Both jobs were pioneered by Penny Nii, the first knowledge engineer and Feigenbaum's wife. Extracting knowledge often required several cycles: experts didn't always know exactly what they knew, nor could they articulate it. Seeing their expertise laid out in code or seeing the results of an executed program, they might realize they'd forgotten to mention an important step, mischaracterized the importance of it, or identify any number of other glitches that became apparent only after the program was run.

But once knowledge was successfully extracted and coded, it was an extremely powerful way of solving real-world problems. Dendral's success also came because it solved a relatively narrow and well-defined problem with clear solutions. Although Mycin, the infectious disease-detecting program, often outperformed the Stanford specialists in that task, it too was ahead of its time. Because it couldn't easily be integrated into local area networks, it wasn't useful for a physician on the job.

Knowledge-based systems, as they came to be called, would permeate AI, whether humans jump-started the program's knowledge, or the machine collected and interpreted the knowledge autonomously, as would happen in the early 21st century in machine learning and data science.

Developments in computer technology helped AI's successes in the late 1960s and through the 1970s immensely. Solid-state hardware, telecommunications interfaced with computers, better time-sharing, more sophisticated software generally all made expert systems possible, practical, and then commonplace.

3.

I vividly remember Ed Feigenbaum visiting Carnegie Mellon in the early 1970s and addressing his colleagues about his expert systems research. "Guys, you need to stop fooling around with toy problems," he declared to researchers engaged in chess and speech understanding. It was a nervy challenge to his two great mentors, Newell and Simon, and to Raj Reddy, who, after all, had been hard at work on making computers understand continuous human speech, hardly a toy problem. Yet if Feigenbaum's comment bent noses out of shape, I didn't hear about it.

Feigenbaum was convinced that the scale of AI itself needed expansion. AI was being practiced by a handful of people, and there was no source book. Thus was born *The Handbook of Artificial Intelligence*, an encyclopedia of all that was then known in AI. It was important to the field's growth, and its royalties went to the Heuristic Programming Project at Stanford to support yet more graduate students. After the book made these principles public, researchers all over the world, especially the Japanese, would seize and develop them.

4.

After I left Stanford in 1967, Ed Feigenbaum and I remained good friends, phoning each other (too early for personal email in those days) or dropping in on each other on either coast. A comfortable harmony existed between us, because each of us was working, especially in the late 1960s and early 1970s, on how to reshape the roles of man and woman, husband and wife, inherited from our culture. How could we live a life that was both fulfilling, yet considerate of those we loved?

The path wasn't smooth or obvious. Ed saw me not long after Joe and I first moved to Pittsburgh and later confided that he'd been worried about me. He could see I was already resentful of the long hours Joe spent at Carnegie, but I seemed to have no life of my own. That was the winter I explored Pittsburgh and western Pennsylvania by myself, knowing no one, marooned in an alien landscape. What degree of autonomy I could allow myself? I'd begun writing a novel about TV news. Was it okay not to be home to fix dinner because I was sitting in a TV studio watching a news program being produced? Tradition said my husband was free to do his job at any hours he chose, while I must wait passively to have my time programmed by his schedule. It was puzzling to work out.

In mid-August of 1972, Ed and I met at Stanford and spoke frankly of our friendship and how important it was to each of us. My journal records our very personal exchange. I told him I loved him because he'd known me in bad times and good, and I always felt like he was on my side. Above all, I said, he knew how to listen.

Ed protested, "I'm not an indiscriminate listener. I listen to you because our thought processes are so much alike, and I feel like we have a special understanding because of that. I can talk to you in turn because I never feel as if you're judging me. You understand, you accept, period. I'm always scared before we meet that somehow it won't work, that I won't be able to convey to you that I'm—"

I interrupted. "Me too, because every once in a while it doesn't click, and I feel sad, and empty, and frustrated."

In my journal, I wrote:

> *A magical afternoon, the sun as tangy as club soda, the blue sky and green*

> *of Stanford's trees vivid. I didn't want it to end. When will we see each other again? We know that the friendship would not be as intense if we saw each other regularly, yet we also know that we did see each other daily for years, and our affection and respect were steadfast. I've never felt as warm and affectionate toward him as I do after today.*

When I thought I might write a history of AI, but had doubts whether I could tackle the scientific complexities of the field, Ed stepped in firmly to shore up my self-confidence. Yes, you can, he said; we, your friends, will help you. They did, him chief among them. During the time I wrote that book, I was at my most eager to get out of Pittsburgh, so Ed began inventing jobs for me. The Stanford computer science department might publish a journal, and I could be executive editor. Expert systems research was being commercialized, and Ed was involved with two startups. If I came to Silicon Valley, there'd be a high level job for me at one of those places. What kept me from saying yes was the conviction that I was meant to put my name on the spines of books, not edit other people's words or make the wheels of commerce turn.

So we made do with phone calls. We both loved music; we'd often tell each other about new music we'd heard, wanted to share. Ed was singing in the Stanford Chorus. Years later, he heard I too was singing (though the American Songbook, around a piano in midtown Manhattan) and teased me: what took you so long? We had ambitions to read novels together, exchanging reactions across the continent, and I think we did read *One Hundred Years of Solitude* together. In any case, we'd send each other titles of books that we thought the other might like.

Our friendship has been one of the great blessings of my life.

Raj Reddy and the Dawn of Machine Learning

1.

Raj Reddy had been a gaunt, large-eyed graduate student when I'd met him at the Stanford Artificial Intelligence Lab in the mid-1960s. He was soon to be one of the two first Stanford PhDs in computer science, another member of the second generation of AI researchers. By the time Joe and I went to Carnegie Mellon in 1971, Reddy was on the computer science faculty and was beginning to look less famished, if not yet fighting weight. He'd always had an easy, wide smile, and now it animated a fuller, quite handsome café au lait face, bright eyes, and a small moustache. Marriage to Anu had clearly been good for him, and they were the happy parents of two little girls. They wanted to raise them as much Indian as American, traveling to India every year during the long summer holidays, but Reddy would sigh to me a few years later, "I know I won't be arranging their marriages. They're American. They'll marry who they want."

At Carnegie Mellon, Reddy was leading a project to construct a computer program that could understand continuous human speech. The difficulties were enormous. Whereas Roman letter text conveniently puts spaces between words, periods at the end of sentences, and indentations to signal a new paragraph, speakers do not. Moreover, written words don't have distortions from intonation, hesitation, or background noise that spoken words do. "But it never occurred to me," he once said with a brilliant smile, "that it couldn't be done. That may be exactly what's needed for anybody who wants to go into this field, namely, blind optimism with no reasonable basis for it."

Reddy was born in 1937 in a rural Indian village—Katoor, Andhra Pradesh—unchanged for centuries. When his father went to the astrologer, as he'd gone for each of his sons, the astrologer raised a warning finger. "For this one, make every sacrifice. Send him to school. This will be the one." Reddy was sent to school and learned his letters scratching the dirt with a twig. From there, he eventually went on to engineering school and then for a masters to the University of New South Wales in Australia. After working in Sydney for a year or so, he went to Stanford to study with John McCarthy.

Reddy had intended to study the solution of large numerical problems by computer but was quickly caught up in AI. For a class project, he proposed a speech-understanding program to McCarthy, who said it was good idea, but after a few germane suggestions, left Reddy to himself. "It didn't bother me," Reddy said, "because I was quite happy to go do what I wanted. But some others who wanted to work with John needed a lot more help. They didn't get it because John doesn't operate that way." That early class project led to decades of challenging research. (McCorduck, 1979)

Programs existed that recognized distinctly separate, clearly spoken words. But the problem of recognizing words in continuous speech? And then *understanding* what those continuous utterances might mean? This proposition was much more difficult. In 1973, a committee of prominent AI researchers, chaired by Allen Newell, studied the problems of speech understanding for the Defense Advanced Research Projects Agency (DARPA). The group agreed it was difficult but worth a try and noted an interesting paradox—in spontaneous spoken communication, people seemed limited only by how fast they could think, whereas in writing, the opposite was true: people couldn't write as fast as they could think. Speech was the primary, normal communication between humans, who also wanted to talk with their computers. To figure out how computers could understand speech would be a difficult but fundamental problem to crack.

2.

Reddy soon grasped that the issues in speech understanding were central to AI generally—the balance between an immense number of facts and far fewer general techniques for making sense of them. To understand continuous speech, many different and nearly unrelated kinds of knowledge are needed (semantic, syntactic, pragmatic, lexical, phonemic, phonetic, and so on). How many pieces of knowledge did a person use to decode an utterance? How did the listener decide which kinds of knowledge were more important than others? How did the listener decide he'd finally understood the utterance? For that matter, what *was* understanding?

Reddy and his team handled the pieces of necessary but unrelated knowledge by constructing a system that simultaneously allowed independent knowledge sources to offer hypotheses about the meaning of an utterance. It was as if each of these hypotheses from different knowledge sources were scribbled temporarily on the same blackboard, where other knowledge sources could see and check them and also generate their own hypotheses. The control system allowed different bits of knowledge to be linked from the simplest to the highest-level. “The analogy I use is a Russian, a German, a French, and a British engineer all coming together to design an airplane. Each one of them is an expert in a different aspect of aircraft design. They don’t speak the same language, but they write their solutions on a blackboard, which others can use without understanding how that solution was arrived at.” Reddy emailed me.

Understanding? What was it? Again, Reddy and his team chose behaviorist measures, six different ways to identify that understanding had taken place. Some were straightforward, like giving the right answers to questions, paraphrasing a paragraph, or drawing inferences from it. Some were less so, like translating an utterance from one language to another or predicting what the person might say next. These aspects of understanding work on different levels. As we’ve always suspected, understanding can be deep or less deep. How deep, everyone wondered, did understanding have to be in order to be useful? (Not very, as we’ll see.)

In years to come, the blackboard model would be fundamental to all commercial speech-understanding systems. The model would be adopted across much of AI as a way of coordinating multiple knowledge sources to arrive at a plausible answer to a problem. Hearsay, Reddy’s program, also was one of the first programs to use probability as a measure of belief (for example, the program could determine a word is probably *fix,* not *kicks*). Now, this statistical technique underlies much of modern AI.

One September morning in 1976, I sat in on an early demonstration of Hearsay. Next to me was Herb Simon and I told him about a meeting I'd been to that summer in Los Alamos, where I'd met the elderly German engineer Konrad Zuse, builder of the Z3, the first fully functional, program-controlled electromechanical computer, which he constructed in 1941 in the living room of his parents' Berlin apartment. Zuse understood at once that his machine could process symbols as well as numbers: he'd invented a programming language called *Plankalkül*, which allowed him to imagine chess playing and other intelligent applications. He'd also expressed some wariness to me about AI. "Playing with fire," he said in his heavy German accent. Simon in turn told me about his and Dorothea's summer, eating and drinking their way through southern France.

The Hearsay demo began and hushed us. A sentence was spoken aloud, so we'd know what Hearsay was hearing, and how it responded.

Hearsay responded by crashing at once. We smiled. This wasn't—isn't—unusual with the first few runs of any computer program. (It was the major common-sense argument against President Ronald Reagan's pet defense program known popularly as "Star Wars," which must work perfectly the first time.) As the wizards re-tuned, Simon and I chatted some more, easily understanding each other, gesturing to fill in, elaborate on, and shade meanings, using incomplete sentences, stopping to laugh, as we always did. Again, speech is the original human communication, and long precedes writing. Yet it was vexingly difficult to teach computers how to listen and understand.

After a while, Hearsay resumed. This time, success. What we heard (and saw) would soon yield a system—the first of several generations of systems including Dragon, Harpy, and Sphinx I and II, each redesigned for faster search—that won a DARPA award. For example, Harpy was the first system to understand, with less than ten percent error, continuous speech in anything like real time. If Harpy wasn't entirely sure what it heard, it could make a best guess. In that, it was very human-like. True, its vocabulary was only a thousand words and confined to a narrow domain, but Harpy was a serious beginning. What it also showed is that knowledge is mainly dynamic, not static.

3.

In *Machines Who Think* I wrote:

> The symbols that stand for knowledge are entities with a functional property. Symbols can be created; they lead to information; they can be reordered, deleted, and replaced. All this is seen explicitly in computer programs, but also seems to describe human information processing too. Understanding is the application—efficient, appropriate, sometimes unexpected—of this procedural information to a situation, the recognition of similarities to old situations and dissimilarities to new ones, and the ability to choose between doing the small repairs, or debugging, and changing the whole system.

Harpy, a successor to Hearsay, was, among other things, the stark recognition of how important context is to understanding. Yes, philosophers had long asserted that context mattered, but philosophers asserted endlessly with not much to show for it but assertions. They might be correct; those assertions might even match our intuitions. But assertions aren't proof. Neither are intutions. Harpy was proof.

As significant as Harpy was, the local Pittsburgh newspapers yawned. The national papers were oblivious. (No one expected TV to pay attention.) But when Harpy won a DARPA award from the Defense Department, John McCarthy thought it was worth making a fuss about and informed *The MercuryNews*, Silicon Valley's hometown newspaper in San Jose, which published an excited story. In Silicon Valley, people got the significance very well.[1] Reddy's graduate students would take the work ever further, including a young Taiwanese named Kai-Fu Lee, who designed the first speaker-independent continuous speech recognition program (and will figure in our story later).

A computer program demonstrably doing its intelligent stuff, however elementary, was thrilling to me. What, really, was understanding? What mattered in knowledge representation, either in computers or even our minds? I chewed on these issues with endless pleasure. I arranged lunchtime meetings with philosophers (it was said that the University of Pittsburgh had infamously bought the entire Yale philosophy department at some point, and Pitt was strong in matters of epistemology and philosophy of science). This was the stuff of my days, and I loved it.

4.

Even in a department of stunning visionaries, Raj Reddy stood out. When he saw the first personal computer with a graphical interface controlled by a mouse (Xerox PARC's Alto machine), he knew that

[1] The next journal to take an interest was The National Enquirer, then a supermarket tabloid of the ridiculous, the spurious, and the lascivious. Reddy didn't want to talk to them, but they threatened to arrive in Pittsburgh and force their way in. Joe got on the phone to them and said, as department chairman, he'd be delighted to talk to them. What were they most interested in? The significance of the new Kung-Traub algorithm? Parallel algorithms and complexity?

the computer science department at Carnegie Mellon must have one for everyone in the department, about a hundred. Funders just laughed. The head of Xerox PARC suggested maybe ten? Raj began to raise money for CMU's own equivalent to the Alto. Dan Siewiorek, an eminent software designer, recalls (Troyer, 2014):

> Raj was like the Wild West. Anything conceivable was possible. He could just go off and do anything. Have you heard of the 'half-Raj' and the 'full-Raj'? The half-Raj is when Raj says, 'Dan, I'd like to talk to you,' and you know it's going to be very interesting. When you get the full-Raj, he puts his arm around you and you're going to be totally involved in a grand adventure.

After Joe and I left Pittsburgh, Reddy founded the Robotics Institute at Carnegie Mellon (its building known informally as Raj Mahal). The Institute was independent of the computer science department and independently funded. It caused uneasiness among some of the faculty—would the high standards expected of Carnegie Mellon students in computer science be maintained? Yes, the Institute jumped to prominence immediately and sustained that prominence, eventually bringing to Pittsburgh firms eager to industrialize that technology, such as Uber. Reddy himself was especially interested in robots that responded to—understood—voice commands.

But then Reddy was interested in so much. Robotics, of course, especially robots that can see, hear, speak, move; interested in language generally, and in specific languages; in human-computer interaction; in machine learning; in software research. He put enormous energy into national and international science and technology policy. For example, he co-chaired the U.S. President Bill Clinton's Information Technology Advisory Committee and was Chief Scientist of the Centre Mondial Informatique et Ressources

Humaines in Paris. He was key to three or four major projects to help bring information technology to the developing world.

After Joe and I left Pittsburgh, I sometimes returned to visit, and the Reddys would always invite me to dinner. They knew I loved Indian food; I hope they liked my company. By now they lived in a big house in Shadyside, just off Fifth Avenue. The house was always full of relatives—cousins, nephews, sons of friends—whom Reddy had brought from India to the United States to give them the same opportunity to study that he'd had. You never knew who'd be sitting down to a sumptuous dinner, but you could be sure it would be great fun.

I came away from these dinners unaware I was trailing a cloud of Indian spices—turmeric, coriander, cumin, cardamom—into the home of my hosts, Lois and David Fowler, who dined on unadorned New England fare. Lois only told me much later after she and David had been invited to the Reddys' for dinner and came home in the same cloud. "We always wondered," she said. "Now we know."

One mild May evening in 1981, I was at the Reddys' table yet again. We ate Anu's wonderful Indian food; we drank a superb bottle or two of Chateauneuf du Pape (Reddy's stint as the chief scientist of the Centre Mondial in Paris had certainly enhanced dinner wine selections). The children went off to do their homework; the cousins, nephews, and uncles drifted out.

We three lingered over our wine. The Reddys liked the lights to blaze, and I wondered if this was a defiant response to childhoods spent by smoky lantern light. All of us were in our forties now, and the blaze was unforgiving of the deeper creases in our faces, the darker shadows under our eyes, the start of graying.

Our mood turned dreamy. We reminisced about days at Stanford fifteen years earlier. Anu urged me to come with her to India and talked about what we might do if I visited her village. We laughed about womanly things, and some years later, when I was on a bus in Tokyo to the shrines and temples of Nikko, surrounded by Indian tourists, I took in the women in their plastic barrettes, their synthetic saris, and heard Anu's voice in my ear: "Oh, Pom, those artificial saris are so shobby; you wouldn't be caught dead in one." I blew her a kiss across continents.

Reddy that evening told me the story of his father and the astrologer. He added impatiently and sadly how wrong the astrologer had been. "Each of my brothers, plowing the fields, is as smart as I am. I got the opportunities. They didn't."

For this was what truly drove Reddy. It wasn't just an eagerness to conquer the next scientific or technological problem, although there was much of that. It was his own life. Unlike some Westerners (and, for that matter, Mahatma Gandhi) who entertained fantasies of a prelapsarian village life, uncorrupted by any technology more complicated than the spinning wheel, Reddy had grown up in such a village. He knew that people in such villages—not just in India, but in China, in Africa, in the Americas, the Arab world—were trapped inside a corrosive, deadly ignorance. They were prey to demagogues and anyone else who wanted steal from them, cheat or manipulate them; they were prey to diseases they needn't suffer; they were vulnerable to, and often overwhelmed by, customs and traditions that smothered the human spirit. In those villages were young Raj Reddys, hungry for the world's knowledge—maybe to make use of it, but above all, to taste the joy of knowing it.

Thus Reddy believed every step he took toward improving and distributing information technology was sacred. AI wasn't about replacing humans, pushing them out of jobs, making them superfluous. It was about releasing humans, not only from literally backbreaking, knee-pulverizing, mind-numbing work, but from oppressions of every kind.

Driven always by the realities of learning to write in the dirt, at the expense of his equally gifted brothers, Reddy wasn't just a scientist of great distinction (the Legion of Honor from François Mitterrand in 1984; the Turing Award in 1994, shared with Ed Feigenbaum; the Padma Bhushan award from the president of India in 2001; The Okawa Prize in 2004; the Honda Prize in 2005; the Vannevar Bush Award in 2006; a slew of honorary degrees; and in 2014 a fellow of the National Academy of Inventors). He also became an activist in international computer education.

When I went to West Africa in 1982 to look at computer education projects, I went with a list of phone numbers from Reddy. Those numbers connected me with West African field experts of the Centre Mondial, the early 1980s French effort meant to bring computing to its former colonies. I'd write about that wonderful experience in *The Universal Machine*.

What I didn't know was that Reddy had made sure I'd not only be welcomed, but looked out for. The phone numbers reached to the top of the West African power network. When I returned and thanked him, he smiled, shrugged deprecatingly. "I wanted to make sure you were always okay. Some of those places can be pretty rough." I hadn't even known I'd run any risks.

Reddy helped found the Universal Digital Library. Although the library was planned to hold a million books, by 2007, it already had a million and a half scanned volumes, readable for free over the Internet. He wanted to see every book possible made available to everyone on earth. “Even pornography?” I teased. “Sure,” he retorted. “No censorship. I mean *every* book.”

Google Book Search, the Gutenberg Project, and the Internet Archive book scanning projects eventually replaced this effort, but Reddy was happy with that—so long as the books were up, available, readable. He considered his own project a proof of concept and was glad for others to take it over.

His work for his native India would be indefatigable. He was a founder of the Rajiv Gandhi University of Knowledge Technologies, intended to serve the educational needs of gifted rural youth; he served on boards and governing councils of other Indian universities; he was active in a network of elementary schools that served the poorest of the poor.

5.

Meanwhile at home, when computer science in all its aspects at Carnegie Mellon had outgrown what any single department could encompass, a School of Computer Science was founded (largely engineered by Allen Newell) and Reddy served as its dean from 1991 to 1999. There, among other things, he helped to establish a Department of Machine Learning. It was a sweet tribute, I thought, to the original Hearsay.

In 2014, Reddy mused to me about early AI research. He was distinctly skeptical about its original biases:

> From Turing on, intelligence was really only human behavior in an

> educated world. Did that mean all those illiterate billions in the world weren't intelligent? Of course not. Any society that can invent writing and zero must be called intelligent.

Babies at birth, he went on, are already pretty smart—they soon have the capacity for language, acquired in the womb. Large numbers of connections between the motor and the visual cortex fire up in early babyhood. "In short, we need to think of the different levels of behavior that comprise intelligence. You can't get to four-year-old intelligence until you've done three-year-old intelligence."

He calls his latest project Guardian Angel technology, a way of getting the right information to the right people at the right time. An intelligent agent can scan vast amounts of information on behalf of its "ward," learn what its ward can't know, decide what's important, what's relevant, what knowledge might protect that ward, and whisper into its ward's ear. It won't be supernatural—it can't predict the unpredictable.

The biggest problem will be how the agent decides importance. You don't want continuous, boring, day-to-day stuff. You want only warnings about possible misfortunes. He envisions this for every man, woman, and child on the planet, using cheap wearable computing. "We can do this," he says confidently. Microsoft Research has a similar set of programs under development for down-to-earth reasons. The time is right.[2]

Jason Hong, one of Reddy's colleages at CMU, envisions a variation of the Guardian Angel called Maslow (named for the thinker

[2] As for improvements in voice understanding, key members of the team that originally designed Siri, your voice pal on your smartphone, have moved to a start-up called Viv, where they envision a consumer-friendly personal assistant you can talk to, connected to a global brain in the cloud that will respond in depth. Hackers smack their lips.

Abraham Maslow, who proposed that humans have a pyramid of basic needs to be met). Maslow the program is a set of personalized agents that "can help us find, set, and meet hard goals in meaningful ways that we choose. Think of it as a cross between a lifelong coach, a caring uncle, and an honest and supportive friend. . . . Healthcare is a clear case where humanity needs significant help in achieving hard goals" (Hong 2015). Hong continues by giving examples of how Maslow might help meet those goals. What if we decide we'd like to be more green? Or wanted to learn painting? "Maslow might even help us find compelling new goals to set for ourselves in forms that are fun and engaging. . . . Maslow could incorporate deep ideas from psychology to help motivate us and sustain changes in our behavior, all the while ensuring that the interventions we get are commensurate with our ability and level of motivation" And Hong gives reasons why Maslow is entirely feasible within fifty years.

In September 2016, Reddy gave the keynote address at the IBM Cognitive Colloquium. He displayed the desired attributes of an intelligent assistant, a list he'd long ago drawn up with Allen Newell: it should learn from experience; exhibit goal-directed behavior; exploit vast amounts of knowledge; tolerate errorful, unexpected, and possibly unknown input; use symbols and abstractions; communicate using natural language; respond in human reaction time (milliseconds). Kenneth Forbus, the distinguished AI researcher who posted this in social media, reminded us that this "tells us how far we have to go, compared to where we are now, despite amazing progress."

Natural language processing has come a very long way since Reddy's pioneering efforts. Low-cost household gadgets you talk to in natural language, like Siri, Alexa, Google Assistant, and Echo, are growing in popularity. In 2018 in San Francisco, two award-winning college debaters, Noa Ovadia and Dan Zafrir, debated an IBM program called Project Debater on the topic "We should subsidize space exploration," followed by "We should increase the use of telemedicine." The audience declared the outcomes a draw but thought Project Debater (meant to exhibit IBM's ability to consult very large data sets, including news articles, and convert that information into flowing, spoken prose) conveyed more convincing information than its human opponents but was less persuasive as a rhetorician. IBM envisions Project Debater as an assistant to human decision-makers, supplying them with evidence-based arguments for one position or another in the midst of conflicting opinions (Solon, 2018).[3]

But as Reddy had imagined more than half a century ago, we're talking to our machines—and they're talking back.

One of Reddy's graduate students, Kai-Fu Lee, Taiwanese-born, now Beijing-based, who got his PhD for the first speaker-independent program to recognize continuous human speech, would disturb the world in 2018 with a popular book about the coming confrontation between the world's two AI superpowers, the United States and China. But that was to come.

This book has focused so far on how the early science of artificial intelligence began, prevailing against a major storm of scientific scorn generated by the peers of these early AI scientists. As late as 1975, when the field of computer science was preparing a progress

[3] Emphasis on this new direction for Watson might be because, despite estimated billions in investment, IBM's hopes to bring AI to medical care were in some ways disappointing. We know that sooner or later, AI will shape medical care, but maybe Watson won't necessarily be in the forefront.

report on itself for the National Science Foundation, the committee planned to omit any mention whatsoever of artificial intelligence. Only when one of the most distinguished scholars in the field, Donald Knuth, insisted that AI be included, did the establishment relent. Meanwhile, the private sniping was vicious. I know. I heard it.

Despite AI's dubious early reputation, it began at once to elucidate the nature of human intelligence and would push scientists to look at intelligence in other species.

But I also became involved in AI and its fate, as we'll see in the next section.

Part Three: Culture Clash

You were meant to travel
From pillar to post to find yourself.
But the road has its own mind;
You can't tell it where to go.
You arrange tales, just tales.

—Chester Johnson, "Á Pied Through Arkansas"

Whiplashed by the Manichean Struggle Between the Two Cultures

1.

I began writing *Machines Who Think* in the mid-1970s even as I was interviewing the pioneer scientists, and my agent distributed a book proposal. No one in New York publishing knew what artificial intelligence was. Once the topic was defined for them, they failed to see why it was important. "We've already published a book on computers," one editor replied to my proposal in 1974. "Too bad it's too late," another said cryptically, as he too rejected the idea. You can't really blame them. The whole idea was preposterous—Machines? *Thinking*?

These Delphic utterances discouraged my agent and deeply unnerved me. Was it all my imagination? Andrei Yershov at the Doctor program terminal, opening his heart to a nonjudgmental computer interrogator? The Logic Theorist inventing a better proof to a theorem than the brilliancies of Alfred North Whitehead and Bertrand Russell? The robot arms swinging autonomously within their protective plexiglass cages at Stanford and MIT? A waist-high Alberich, shake, rattle, and rolling around the halls of the Stanford

Research Institute? It was called Shakey, as if we'd returned to some long-ago, less tactful age when humans named each other bluntly—Fatso, Shorty, Gimpy, Cruikshanks. Shakey wobbled its way around, evading walls and people, veering to a plug when juice ran low.

Machines all around me at Carnegie Mellon, MIT, Stanford, and Stanford Research Institute were busily performing all sorts of tasks: playing chess, playing poker (though not as well as they would by 2017, when two different programs beat professional human champions at Texas Hold 'em).

At Carnegie Mellon, the Cheese Co-Op program took monthly orders for cheeses unavailable at the Giant Eagle, the local supermarket. The program then calculated how much cheese of various kinds must be purchased at the wholesale cheese market in the Allegheny River's Strip District. The program did a further calculation of how to cut up wheels and wedges optimally to meet each family's order. The project's popularity soon meant hundreds of pounds of cheeses were delivered to the campus. (Joe said silent prayers that the funding agencies didn't discover that particular AI student lark. "The first e-commerce," Raj Reddy joked at a symposium in 2015.)

In March 1975, after describing my computational life to a skeptical group of high school science and mathematics teachers, I realized how far I'd come from my original agnosticism, how persuaded I was by AI. "I'm no longer a disinterested party," I confessed to my journal.

2.

In the mid-1970s I lived in a world that would be commonplace three decades later—a computer-saturated environment; email; AIs both visible and invisible, audible and inaudible, and all of them partial intelligences (which threw off outsiders who still held that intelligence was—or it wasn't). But few others shared such a life, so no wonder outsiders found it difficult to believe on any level. Yet I claimed I was writing science, not science fiction.

Maybe because this world was everyday to me, I couldn't write *Pop! Wow! Zowee!* I thought the facts supplied their own pop-wow-zowee without rhetorical flourishes. I hoped to tell the story straightforwardly, with an earnestness that, okay, maybe skirted scholarly tedium. I failed to see how improbable the AI world seemed to anyone outside the field.

My agent and I parted, and I found another who was more enthusiastic. But the rejections kept coming. I was so demoralized after nearly two years of trying to sell the book that Joe offered to intercept the rejections for me. If I had to read these responses, most of them ignorant, not all of them courteous, I wouldn't have the fortitude to go on. Halfway through the manuscript, I wondered if I should even bother finishing. The work was hard; nobody wanted, much less appreciated it. Years wasted. The Two Cultures clashed again, and as far as New York publishing was concerned, I'd picked the wrong side. They could hardly conceive there was another side to pick.

From my journal, November 28, 1976:

> *I could use someone to talk to just now, to explain, or figure out, why I*

> *feel under such a strain, so dissatisfied, so blue. Partly it has to do with my work. I'm 36, and my work will never be better (I think) yet it's essentially being ignored. I'm sick to death of the AI project—it takes much longer than I thought—yet wouldn't dream of not finishing, even though I resent it taking the best of me and my time just now. And even with this sure-fire, can't-miss project, publishers rise on the horizon like mirages, only to dissolve when I get close. So my professional life seems a shambles, my personal life all screwed up on account of first, my professional life, and second, that I've come here to Pittsburgh with Joe on account of* his *professional life, and I wouldn't regret that, to speak of, except for his own deep doubts about whether he made the right choice, putting professional before personal.*
>
> *So I feel like a rat in a maze, with no exit except to grow old. Yet that passivity violates every sense I've had of myself since I was 26—ten years now, nearly a quarter of my life. Just writing about it makes me angry at myself, and anxious to grab my life and shake it into shape. But sometimes I feel like one of those Beckett characters: "I can't go on, I must go on, I can't go on, I'll go on."*

I wanted the AI enterprise—theirs and mine—to succeed, though I had only vague ideas of what success would mean and when it might arrive. If the field had seemed mysterious, it was no longer. Difficult, yes. Scientists I interviewed and wrote about worked tirelessly to gain centimeters of progress. From the beginnings of history, humans had imagined creating intelligence outside the human cranium, and later, hoped to understand the human mind in a scientific way. With AI, they were trying to do both.

Finally, after more than thirty rejections (outdoing *Catch-22* by several), Joe made a couple of phone calls and put my agent in touch with W. H. Freeman, the book-publishing arm of *Scientific American*. An editor there, Peter Renz, was intrigued. Renz and I had several meetings, and he seemed enthusiastic.

3.

But between the day Renz said he'd offer me a contract and its arrival, some four months elapsed. Fate intervened melodramatically. In a routine medical visit, a deep internal tumor thought benign had tripled in size within six months. Maybe not benign after all. A second opinion wasn't reassuring. I'd need surgery to tell for sure. At thirty-six, I came face to face with mortality.

What would you do if you had only a year to live?

A game we'd played as young adults, but now it wasn't theoretical. I faced the question for real and knew instantly. I must get out of a place that made me miserable and back to warmth, sunshine, beauty, and my family. I had to stop deferring my own desires until the perfect job for Joe came along.

The tumor proved benign after all, but the crisis focused me. Even as I signed the contract with W. H. Freeman—May 22, 1977—and was still recovering from surgery, my path ahead was clear: finish the book and get out of Pittsburgh. It didn't have to be in that order.

Lois Fowler, a professor of English at Carnegie Mellon, and my closest friend in Pittsburgh; the novelist Mark Harris, my colleague and friend in the English department at Pitt; my friend Herb Simon; my husband Joe; each urged me to stay another year to go through the tenure process. It would be an asset on my resume, they said. My department chair had already encouraged me to start the process, and no one anticipated problems—the Pitt English department's tenured faculty then comprised scholars who'd published an article or two, or none. I was under contract for a third book. My student evaluations were fine. "All right," I said, "one more year."

But the surgery had distressed me at levels I hardly understood. Depression turned into desperation. Doubtful, but knowing I couldn't go on as I was, I took my friend Lois's advice and began therapy with a psychiatrist, Charlotte Babcock, who might at least see me through this further purgatorial year.

And I wrote. When I had what I thought was a decent draft, I gave copies to Newell, Simon, and Marvin Minsky to check for technical details. From Newell, I received a four-page single-spaced commentary, very perspicacious, I noted in my journal. He followed that a few days later with more comments, speaking astutely about my persona in the book. On September 28, 1977, I also received two pages of comments from Simon. "*They both care very much that this book be done right, and have been monumentally supportive,"* I wrote in my journal.

What perturbed Simon was whether I'd got right the impact of the Logic Theorist, the first example of a symbolic thinking machine, and the information-processing model that he and Newell had presented at the Dartmouth meeting. He didn't think so.

October 7, 1977:

> *Worked, tussled later with Herb about whether the information-processing model had really been seen as important at the time. We went over Minsky's early version of "Steps Toward Artificial Intelligence," and it seems not. Gamely, Herb conceded the point. But I'll rewrite to show the paradigm shift was heralded at Dartmouth, if not recognized.*

October 8, 1977:

> *Worked the day revising the Dartmouth chapter, and finally got a hypothesis that fits the data and even sounds plausible to Herb, whom I called in the early*

> *evening to check it out with. The information-processing model seemed relevant only to psychologists, and most AI types didn't feel compelled to make AI resemble natural intelligence, necessarily. When I called Herb, he told me he'd spent the day digging around in his own archives, trying to solve the problem. Also, as I was going over the comments he'd made in the margins, I saw the first vulgarity I've ever known him to use: SHIT! he'd scrawled at something Lotfi Zadeh had said. I laughed at that, and plenty of other comments he'd made. That marked-up version of my ms will go into the same archives as the tapes and transcripts. What a treasure he is.*

October 10, 1977:

> *Sat down to write a note to Marvin on the ARPANET* [early email] *and discovered the line-by-line comments from Allen on my first two chapters running on and on and… Again, my heart sank. But they're not so much arguments as musings, which is even better. And reminders of promissory notes I give the reader to be paid off in due time. I was very, very pleased to get such detailed, loving attention.*

Minsky sent email saying "Just lovely." I wondered if he'd even read it, but later, he commented on the manuscript in detail.

November 4, 1977:

> *A much greater sense today of being "finished" with the book, a sense of closure, of tying up loose ends. I'm pleased with the last chapter, despite its windiness. I've tried to capture the sense of it all being tentative, a prologue to the really big stuff.*

I delivered a finished manuscript to W. H. Freeman during the Christmas holidays of 1977, relieved to be finished at last.

4.

But that spring of 1978, the tenure process did not go well. My department recommended me only by a ratio of two-to-one, unhappy with how difficult it was to categorize me. Was I a novelist, as my first two books said, or a nonfiction writer, as this new book seemed to say? The doubters couldn't understand where I was going as a writer (as if I knew, as if writers methodically laid out a life plan) while the writers on the faculty argued that my commitment was to writing itself.

I wasn't blameless. In 1976 I'd written a provocative piece that appeared in *The Chronicle of Higher Education* called (not by me but by the editor) "An Introduction to the Humanities with Dr. Ptolemy." There I'd argued that the humanities were demoralized, and their exaltation of the human species alone—human chauvinism, both Lewis Thomas and Carl Sagan had called it—was as out-of-date for a guide to being human as the Ptolemaic system was for navigating the high seas.

I referred to a long exchange between C. P. Snow (remember The Two Cultures?) and F. R. Leavis, an influential British critic at Cambridge. Leavis had written that the world's problems were to be solved by "mankind . . . in full intelligent possession of its full humanity . . . something with the livingness of the deepest vital instinct; as intelligence, a power—rooted, strong in experience, and suprememly human—of creative response to the new challenges of the time. . . " (Leavis, 2013). I said I'd read this aloud to my students, and asked them what it meant. They laughed. I confessed that I'd laughed with them. "If we'd written that in one of our papers—" one student began. But Leavis was one of the greatest of the humanists, defending the humanities. "The humanities are demoralized because they are no longer adequate for us in the world as it is," I wrote. "We are hard up for a Copernican revolution that will take man from the center of the universe and put him somewhere more appropriate." (McCorduck, 1976)

This offended my humanities colleagues, and the lukewarm tenure vote was to let me know.

How did I feel? "Like I've been gangbanged by dwarfs," I said to somebody at the mailboxes, which went viral around the department. "Like I've done right and been deeply wronged,"I wrote in my journal, which went on to note:

> *In a way it's laughable, and in a way it's terrible. I told Mary* [the department chair] *that what I resent most of all is being penalized for stretching, taking risks; that I've worked very hard and stretched far, and it's not so perceived by my colleagues. I saw Mark Harris briefly. "Don't take it personally," he said. "Other things are going on." But how else should I take it?*

Later, a male colleague, here nameless out of courtesy, said:

> *There's no way you can prove it; there's no way you could make a case, but I believe you were discriminated against because you were a very threatening woman to some of the men who cast votes. I picked it up from body language, from the strange nature of the comments that were sometimes made." I said I'd suspected but hadn't wanted to believe it. Told Lois, who said Mark said he was absolutely baffled by the comments and the vote. That might account for it.*

It was piquant to hear later that this sorrowful male colleague had helped lead the charge against me.

More nonsense leaked out. Some of my colleagues dismissed Herb Simon's letter on my behalf as merely one of my husband's "business associates." Others cried that *Machines Who Think* showed I'd "sold out to the machines."

Sold out to the machines? Sloan Wilson, author of that emblematic 1950s novel about selling out, *The Man in the Gray Flannel Suit,* wrote: "Selling out was doing something you did not want to do for a good deal more money than you got for doing what you loved to do" (Halberstam, 2012).

No. I was doing exactly what I loved and getting nothing but grief. Having gone through the same arguments for two years with publishers, how could I be surprised? *I, Simplicissima*.

The official notice from the dean said that the university preferred to wait with the tenure decision until they could see how *Machines Who Think* was received. That struck me as a dismal acknowledgment of intellectual poverty—just what I'd complained about in "An Introduction to the Humanities with Dr. Ptolemy."

I'll never know my greater offense: being a strong feminist or being eager to move on from the past, compelled by what I saw as the future. It didn't really matter. My colleagues had released me. I exhaled. I saw again how far I'd moved from other humanities scholars. The distance across Panther Hollow between the University of Pittsburgh's English department and Carnegie Mellon's computer science department yawned as wide and deep as the Two Cultures could possibly be from each other.

A Turning Point

1.

As this tenure spectacle played onstage, my sister Sandra called me on April 24, 1978, to say our father had been diagnosed with acute leukemia. He had only months to live. My grief was immediate, profound, striated with grief for myself. I loved him and could hardly imagine life without him on this planet.

My two loving siblings, twins Sandra and John, were raising young families and lived near my parents in Contra Costa County, east of San Francisco Bay. My father didn't need me especially. But I needed him. I wanted to say goodbye to him properly, to thank him for being so ambitious and courageous that he'd left a secure job as a police officer in Liverpool and heroically brought to the United States a wife and three children under six, determined on better things. I wanted him to know I loved him deeply for his wild, ravenously curious, uncultivated mind, his exceptional sense of humor—he made us laugh even as he was dying just as he always had when he was alive and vital. I wanted to thank him for turning me into the ferocious woman warrior I'd become (even if he'd done it inadvertently with his old-school patriarchal despotism; yet he was proud of me).

Joe was due a sabbatical, so we took the academic year of 1978–1979 in Berkeley. Joe taught on campus, and I buried myself responding to editorial comments on *Machines Who Think*. An anonymous reviewer had undermined my editor's confidence in the book, using the very same devices Mr. Anonymous criticized me for, personal reaction and opinion. Well, yes. That was the point. The book's subtitle was *A Personal Inquiry into the History and Prospects of Artificial Intelligence*.

So my editor and I haggled over the months, this wearing me down almost as much as my father's dying did. I had to remind myself again and again that my name would be on the spine and that every foolish compromise I made out of fatigue or trying to please would appear as mine alone.

On July 25, 1978, *Machines Who Think* went to press at last.

2.

I was determined to stay in the West. I took a job as a technical writer for a subcontractor at Lawrence Livermore Laboratories. I edited safety manuals for projects ended years earlier, "just for the record," I was told. I held my peace on that. I visited my father each evening, a sweet twilight in his life. He loved Joe and wished I'd compromise. "You don't want me to be a bitter old lady," I protested softly. No, he didn't. But he was deeply saddened that this marriage was coming apart. I felt I *was* compromising: I'd be willing to live in Berkeley or Palo Alto (Ed Feigenbaum was inventing jobs for me there) while Joe continued at Carnegie Mellon. Commuter marriages were becoming more common. But Joe wouldn't hear of it.

Our struggle was all embarrassingly public. "Does Pamela know what she's asking Joe to give up?" one of Joe's Berkeley colleagues asked.

I knew he was voicing the consensus. Only my therapist, Charlotte Babcock, had ever asked whether Joe knew how much staying in Pittsburgh was costing me.

Manifestos are cartoons. They make no allowance for the living, breathing humans who comprise couples with individual, sometimes conflicting, desires and hopes for the wellbeing of their partner, for the wellbeing of the union. They make no allowance for the oscillations of ascendancy, first of one partner, then the other, responding to opportunity, to responsibility. They say nothing about love, respect, admiration. I mean the patriarchal manifesto, which dominated human culture for millennia (and in many places still does) that has been encoded into religious and secular law, embedded deep in mores, and assimilated into individual consciousness, shaping everything from science to art to law.

The feminist manifesto was an understandable and justified revolt against that patriarchal manifesto, but it was cartoonish, too, with little nuance or allowance for all those human qualities I've just named. I worried about it, examined myself, and wondered how different I was from the women I knew. They seethed about the unjust advantages of male privilege, muttered quiet confessions of rage to me. Had I been hypocritical to teach my students about feminism, embody it for them to some extent, and yet continue to remain rooted in a city I found unendurable? No. I gave what I did to a man I loved, respected, admired, to what I hoped was the wellbeing of who we were together. Until I couldn't any more.

Josephine Harris, Mark Harris's wife, said to me, "You're just approaching your most creative period, which for a woman, is in her forties. You're going to go off in astonishing directions. Do it—let it happen! Try also to remember these words coming from a woman who wants you to do it for her, in her stead, because she couldn't. Biology makes you make choices too soon. I was married with three children, three hostages to fortune, before I realized I should never have been married. At all."

3.

In October, I returned from the Berkeley sabbatical to Pittsburgh for a few days to gather belongings from our house. My visit coincided with the announcement that Herb Simon had won the Nobel Prize.

From my journal, October 16, 1978:

> *Herb got the Nobel Prize in Economics today! I got the news from Allen, when I called to say hello to him and Noël. Very, very pleased for Herb. I remember our conversation once about the Nobel. He'd found out he'd been nominated, and that raised all sorts of irritations in his mind: he'd rather not have known; it opened up possibilities and aggravations he wanted not to think about.*
>
> *The definition of grace: Herbert Alexander Simon, new Nobel Laureate, found time to call me on this very day because he'd heard I was in town. Could we get together? I was speechless. I wouldn't even have tried to call him—just left a note on his doorstep, saying how delighted I was that he'd received The Prize, that* arête *was seldom so appropriately awarded. We'll try and see each other Wednesday.*

October 18, 1978*:*

> *A half hour with Herb. We hugged—I felt such a rush of joy being with him. He'd had an inkling—the Swedish equivalent of the* Wall Street Journal *had asked for his picture a few weeks ago, as one of the leading contenders; then a former student, now financial reporter for a Stockholm newspaper, called to say he'd spoken with a committee member, who said, "I can't tell you the result, but*

> *I think you'll be very pleased." And what else would be very pleasing except his former professor had won? Thus he called, suggested Herb be awake at 6:00 a.m. the next morning. "It wasn't a problem," Herb laughed. "I could hardly sleep that night."*
>
> *He's so dear and funny. We talked very personally about the effect of the award, that "perhaps it would take the edge off some of his competitiveness." I asked what he meant, and he said now he could stop worrying about who else got The Prize. I had to laugh. He was also glad for the boost it would give his kind of economics* [behavioral] *for youngsters who wanted to go into the field.*[1] *"They couldn't get a dissertation through these guys here," he said, referring to the Friedmanites who now run the CMU economics department. We laughed together rather wickedly over the people who'd be gnashing their teeth—Milton Friedman, for starters. He urged me to come back to Pittsburgh where "good work gets done." I said I missed our afternoon sherries. Yes, he said. Now it was a long walk home.... I left reluctantly because I felt, with his usual grace, he wasn't rushing me, but surely he must have tons to do today. He seemed sorry to say goodbye. Another hug*

I told Noël Newell later that Simon had urged me to come back to Pittsburgh and do good work. "Why would you do that, when you've finally had the guts to break away?" she said, pained.

I returned to California. A few days later, on October 27, 1978, my 38th birthday, we got news that Vera Watson, John McCarthy's second wife, had been killed on her attempt to scale Annapurna. McCarthy made a statement to the press: "She was a woman with a taste for achievement, and I encouraged her to make this ascent." Suffering deep grief myself, I understood how gracious and courageous that statement was.

[1] As I've noted, behavioral economics would thrive in the future and make mincemeat of the rational man assumptions of neoclassical economics with subsequent Nobel Memorial Prizes in the Economic Sciences to Daniel Kahneman and Richard Thaler.

A week later, as my stressed-out mother went to soak in a hot bath, I sat with my brother and sister on a little bench in my parents' bedroom, all of us holding hands. We watched our father die the night of November 2, 1978.

That November was grotesque. A week or so after my father died, a woman was found raped and murdered at the Lafayette Reservoir, a pleasant park near my parents' house where my father jogged daily when he'd been well and where my brother still jogged; where my mother and I were regular walkers. Another few days, and the Jonestown tragedy came. Hundreds of San Franciscans who'd followed a cult leader to a remote place in South America committed mass suicide by drinking poison-laced Kool-Aid he forced on them, which added an acid phrase to the language. After another few days, the mayor of San Francisco, George Moscone, and the gay activist and council member, Harvey Milk, were assassinated by a deranged former policeman, Dan White, who was offended by Milk's gayness, offended by Moscone's liberalism, and who got off later on the infamous Twinkies defense—excessive junk food had disturbed his mental balance.

The world had gone mad. Personal and public tragedies had conflated.

4.

Mercifully, November 1978 was the nadir. In January, Joe told me that he'd been offered a named professorship at Columbia University, with the task of starting a new department of computer science. It wasn't the West, but I was willing to go with him to New York City in March and at least look things over. I'd been happy as a graduate student in New York, and, after all, it was the company town, the self-styled publishing capital of the United States, some claimed the English-speaking world.

Earlier, I'd written an essay for a memorial volume on Gladys Schmitt, the Pittsburgh novelist who was much beloved there. I'd only met her a few times and had a lunch date with her on the very day she died. The book arrived and ended with a short story of hers I'd never seen. It gave me a sudden and peculiar insight into my extreme despair in Pittsburgh.

January 14, 1979:

> *The story has a brief description of a young child in school, and it evoked being in grammar school in the East when we'd first come to the United States. Everything seemed so harsh, so extreme—the cold, the steam heat, the utter misery at home.*
>
> *We five had crowded in with relatives who, after nine months, understandably ran out of patience with our stay. This forced us on other, much more benign relatives for another six months.*
>
> *Then we arrived in California and everything changed abruptly, like the scene in The Secret Garden, where the children step from black-and-white into a Technicolor garden; where Dorothy steps from drab Kansas to glorious Oz. So it was for me. I vividly remember the first hours in California. I was seven, and it's all as clear as yesterday.*
>
> *I must have assimilated all that despair in the face of harsh silencing and extreme stress between ages six and seven. No accident that the Schmitt volume is called* I Could be Mute. *Somehow in the passage of the seasons, the brutality of the environment, in every sense, is always re-triggered by all those seasonal cues in the East. In Pittsburgh, I remember many, many hours of feeling like a child shut up on an endless rainy Sunday afternoon. Very hard to explain to anyone else, but it makes sense—and I sense it as truth.*

Lotfi Zadeh, an old friend, the inventor of fuzzy logic, and always a keen photographer, took a picture of me for the dust jacket of *Machines Who Think*, the last step before the actual book appeared. Joe remarked how much like Hortense Calisher I looked in that picture.

March 2, 1979:

> *A stunning perfomance of Brahms's German Requiem—ever my favorite—in the Stanford Chapel. Ed was singing, of course, but invited us first to dinner at the Faculty Club, along with Jill and Don Knuth, John McCarthy, and Carol Talcott, Penny looking splendid as always. John relaxed and peaceful, though like me, he had a bad time during the second movement of the Requiem. I put my hand in his just to squeeze it, say I understood, but he held on throughout the performance. I was glad.*

5.

On August 27, 1979, two days after we arrived in New York City, the first copy of *Machines Who Think* was in my hands.

> *As Joe and I walked into the lobby of 450 Riverside today, the porter, Norman, was just carrying a special delivery copy of the bound version of MWT. I've been enraptured by it ever since. I've pawed it, read it, been cheered by it. Hell, it's a superb book, that's all. Even the dreariness of dealing with the Chemical Bank couldn't dampen me. I bounced around Morningside Heights all charged up—or even more charged up, since New York has the effect on me that speed has on other people. And wherefore was it glorious? Not because the way was smooth and placid as a southern sea… I am unequivocally proud of myself.*

My favorite private message was from Allen Newell, who wrote three pages, single-spaced. I copied into my journal:

> *"Your book is not merely history but an entry into the intellectual lists, a*

> *way of looking at AI as part of the human enterprise in a big way. I like it better than any other. However, you shouldn't have ignored linguistics and psychology the way you did: it distorts things improperly."*[2]

I commented:

> *Though Allen is mainly praising, and I'm grateful as usual for his attention to detail, I also felt a certain sadness. MWT is not the sweeping scholarly history Allen mentions as an alternative (and that I didn't intend to write) nor is it the popular book that will vault me to renown. I fear I've fallen between two chairs again… The book has disappeared into the gloom of respectable, but unreadable. Unread.*

But my editor called in early February 1980 to say that Philip Morrison had given the book "a very enthusiastic review" in *Scientific American*. "Do you have any idea how unusual this is?" my editor asked. "Phil bends over backwards not to review Freeman's books because of our connection with the magazine, but apparently he really liked *Machines Who Think*."

February 15, 1980:

> *Today I sat in the park and realized I'd reached some great life goal. I'd written a book, and just seen it publicly praised by someone I admire enormously, Philip Morrison, in a place where many people I care about will read its praises. In the March 1980 issue of* Scientific American*, a book I wrote is called delicious, witty, informed, open, rich in direct and candid testimony, offering*

[2] This lapse may explain the unrelenting, and to me, inexplicable hostility from Roger Schank, then in machine linguistics at Yale. One day, my stockbroker invited me to sit in on a due diligence presentation for equities analysts at the University Club in midtown Manhattan. Schank, whose startup was the occasion for this, marched in, saw me, stopped dead, and with his usual courtly presence, yelled, "What the fuck are you doing here?" He actually called my broker afterwards demanding to know why I'd been permitted to come. My broker cleared his throat "Well, she controls a portfolio of several millions." Even a stockbroker's fiction has its uses.

> *a good deal of wise reflection. It is called splendid, judicious, a fine study. So I sat in the park and savored the experience of such lavish praise, such gratifying acknowledgment of all that hard work and grubby obscurity. How did it feel?* It's the best feeling I ever had! *I'm full to bursting! I could've flown unassisted! Let's hear it for fame and praise! It's terrific!*

Forty years later, during a celebration at Carnegie Mellon in April 2018, I heard that the book had inspired and influenced a generation of leaders in the field. These were now senior people. I swallowed a bit to get over the idea that I was *that* old, that the book had been *that* long ago, and smiled happily.

Dissenters

1.

The phone rang before I was up. It was Arno Penzias, the physics Nobel Laureate from Bell Labs. I'd met him when I was sketching a book on computer graphics (the book went nowhere). At some point, Arno and Lillian Schwartz, the computer animation pioneer who used Bell Labs software for her art, and I had a grand time over lunch. Arno and I met socially several times again, and I found him good-natured and likable, if amusingly sure of himself. This morning he wasn't happy. Without so much as taking a breath, he told me for thirty minutes why machines would never, ever think and how deluded and misled I'd been to spend a chunk of my life on such a project. (It may have been this occasion that he told me he thought Herb Simon was arrogant.)

I tried to get a word in edgewise, but the loquacious Arno was not to be gainsaid. His arguments were pseudotechnical or not technical at all, so in my enforced silence, I wondered if his religious beliefs were firing his sermon. He'd once told me that, as a child, he'd been on a train to Poland, part of a massive relocation of Jews of Polish extraction out of Germany. The train was halted by Germany's invasion of Poland. That fortunate stop had eventually saved him from Auschwitz. Since then, he'd had a strong feeling that God had saved him for something special, and thus he was a deeply observant Conservative Jew.

Finally, I pleaded. "Arno, I know you're a married man, so you'll understand. You woke me up, and I haven't even been to the bathroom yet." He roared with laughter and let me go.

2.

The phone rang again on the morning of November 9, 1979. "This is Joe Weizenbaum. I'm calling from Berlin." For forty minutes he picked nits in the newly published *Machines*. This was wrong; that was wrong; I'd misquoted him, misunderstood him. But beyond the nits, I'd represented myself as being neutral, even distant. This was fraudulent: I emerged as a partisan, which I admitted in print.

I'd begun as neutral, I replied, but found myself excited by the audaciousness of such a human project. He objected: I'd thanked three people who'd read the final manuscript, Newell, Simon, and Minsky, which told him whose side I was on. They'd read it for technical content only, I replied, and didn't add that each of them complained about different things. I'd had to resist writing the book each wanted.

Finally, we came to the nub of it all. *Who* had told me that he, Weizenbaum, said he so admired a piece of AI work that he'd have given his right arm to do it? *Who*? I refused to say because it had been said to me in confidence, to confirm my guess that Weizenbaum had been unable to do science and had thus turned to moralizing. Since I'd witnessed the evolution of Weizenbaum's quarrels with AI, I'd written in *Machines Who Think* that there might be correlation between his professional detumescence and his rise as the field's ethical critic. It was plausible and widely believed among his colleagues. I'd disclosed it with regret, I told him, but I believed it. Was it untrue? He didn't reply.

I might not have mentioned this sad little backstory in my book, except his book, *Computer Power and Human Reason* (1976), received remarkable attention, especially from people in the First Culture, who were finally stirring uneasily about computers. Look! Here was one of the Second Culture people, arguing that bad things might happen with the infernal machines.

Yet *Computer Power and Human Reason* seemed to me poorly argued, impressionistic, full of exaggerations and late-age Romanticism, and just plain wrong. It contained long paragraphs, maybe chapters, about the pathetic narrowness of people who imagined they could make computers think.

Who were these spiritually and culturally stunted creatures Weizenbaum was lamenting and lambasting simultaneously? Polyglot Herb Simon, who delighted in music, painting, languages, and literature? Allen Newell, reaching eagerly across, and contributing to, one field after another, but always passionately dedicated to understanding the human mind? Marvin Minsky, fast friends with leading science fiction writers, widely read, thinking more broadly about the brain than most brain specialists, and now composing serious music? John McCarthy, exploring the counter-culture, taking risky political stands, and full of provocative and amusing scenarios he dreamed up, each a parable to illustrate why technology was human salvation, not human menace? Raj Reddy, out to illuminate the most benighted villages of the developing world? Ed Feigenbaum, fearless sailor, avid chorister, and such a lover of literature that, years later, he'd lead long public discussions about the future of the book? Weizenbaum had no right to pass off his fictional stereotypes as authentic portraits.

I was especially offended by his facile arguments that AI could bring about another Holocaust. The most repressive societies then going, I countered, were controlled by ballpoint pens and the gun: China and the Soviet Union. Weizenbaum would certainly prove to be prophetic about AI techniques that corporations and governments use greedily to track us massively, closely, and perhaps unconstitutionally. But this, I'd say now, is a human failure, enabled by, but not the fault of, AI. No science or technology of any significance comes to us unambivalently. Humans must (and we are beginning to) take responsibility in that regard. About the agricultural, industrial, and then the scientific revolutions, nearly everyone agreed that each had its costs, but the benefits outweighed those costs. I believe that about the information revolution and AI, too.

3.

What I didn't realize was that Weizenbaum's book was an early example of AI's Dionysian side, passionate eruptions aimed at stopping the whole enterprise at the same time they soothed and reassured humans in their deep need to be number one. Philosophers, mathematicians, scientists, social critics, literary critics, even public intellectuals would all have a fling. Flawed, neo-Romantic reasoning might repel me personally, but *Computer Power and Human Reasoning* certainly found an audience—it won an award from Computer Scientists for Social Responsibility and launched Joe Weizenbaum on a lifelong career of cheerless lectures about the coming apocalypse.

The book's arguments allowed readers, and later Weizenbaum's listeners, to feel righteous and comforted, without the inconvenience of examining the facts too deeply.

After some further conversation, I said to Weizenbaum that we basically had two different worldviews. I didn't see why plausible arguments couldn't be made for either—my view that life was getting slowly better, or his, that it was getting worse. You could take your choice.

Finally, I asked him what he was doing in the small mill town of Berlin, New Hampshire. Given the staggering costs of overseas calls in those days, it never occurred to me that, as he informed me, he was calling from West Berlin, Germany. (We'd talked for forty minutes on what to me was mostly third-order stuff, and he was spending three dollars a minute to do it.) This led to some discussions about the ambivalence of being a former German Jew in the new Germany. I told him that my husband had escaped Germany by the skin of his teeth in 1939, two months after *Kristallnacht,* and the Holocaust was vivid for us because my husband's parents had lost every member of their immediate families. I'd been born in a rain of bombs that was indifferent to my religion, so long as I was dead. Or at least terrified.

To me, the conversation ended cordially, us agreeing to disagree on whether the world was improving or degenerating. But Joe Weizenbaum was to take revenge.

4.

I'm not sure Hubert Dreyfus's book, *What Computers Can't Do* (1972),[1] was even the first in this series of feel-superior-dear-human

[1] It would be more accurately titled What First-Order Logical Rule-Based Systems

screeds—an example of what, in my own book, I'd called the Wicked Queen syndrome: *Mirror, mirror on the wall, who's the smartest of them all?* Dreyfus's intellectual contributions to the AI debate were finally inconsequential. But he was publicly on the warpath against AI while I was writing *Machines Who Think*, a path he'd brandished his hatchet along since 1962, almost fifteen years earlier. So I felt obliged to interview him.

We met at a panel discussion on the Berkeley campus on May 26, 1976, organized by Lotfi Zadeh, who'd overcome my strong resistance to such spectacles by assuring me I'd have fun. To prepare, I marshaled notes I'd made about the many 19th-century physicians and philosophers who'd averred with pomp and certainty that women could never think nor be permitted to try (grievously ruining the lives of so many of them). In my opening remarks, I drew a little parallel between that and a philosopher who, these days, might be tempted to say that machines could never think. It was meant to make the audience laugh, and it did.

It made Dreyfus furious. His face flushed; he bounced on his chair like the marionette of a demented puppeteer. I noted in my journal that he was vicious, denied statements he'd made, and denied others that no one had made ("I never said women couldn't think!" Who said you had?). I lost count of the number of times he began a sentence with, "That's not what I said; I said…" If I rose to the provocation, I knew I couldn't win. These were the tricks of the rhetorician, and as a philosophy professor at Berkeley, he was a master of rhetoric.

Without Learning Can't Do, write Stuart Russell and Peter Norvig in the third edition of their monumental textbook, Artificial Intelligence: A Modern Approach (2010), although they add wryly that title might not have had the same impact.

Rhetoric is a shadow weapon in science, no matter how convincing it might seem in debate. Results, not rhetoric, are what really count. In this too, I'd moved away from humanists, who'd disagree. But afterwards all the panelists had dinner together cordially, and he agreed to be interviewed for the book I had underway.

To make an anti-AI stance into a busy cottage industry might puzzle me—Dreyfus had been at it since 1962, and his defeat in chess by a computer and other primeval AI tales are in *Machines Who Think*. But to persist so long, the field must have fascinated him. Maybe part of his otherwise unfathomable anger with AI was disappointment that it hadn't succeeded better. Only that, I thought naïvely, could account for his eagerness to attack so passionately, undeterred by any successes the field might have.

Making notes as I went, I drove myself through Dreyfus's book, *What Computers Can't Do*, somewhat outdated by then because computers were now doing some of the things that they were supposed never to do. I wasn't sure I understood it all, but I wasn't sure it was all that clear in his mind, either. Hyphenated phrases, like "being-in-a-situation," presumably adaptations from the German, always make me reach for my peashooter.

After I interviewed Dreyfus on July 21, 1976, I found him likable, "though I surely wouldn't want him jumping all over me with both feet," I noted in my journal, which continued:

> *A dreadfully nervous man—afterwards we walked across the campus to a film, and as he walked and talked, he clutched at his breast regularly, rhythmically, every twenty seconds or so. He was surprised when I asked him why he was so mad at all these AI types. It had never occurred to him to ask himself! After five years of analysis! He hypothesized that it might be that he attacked in*

> *them what he most disliked in himself, an excessive rationality. Can't say I noticed any excess myself. What I did notice was that I'd come to grips with his objections, that I understood them, raised questions about them, which, to my astonishment, he couldn't answer: "Yes, that's weak," "No, I don't have an answer for that."*

His responses to my questions reassured me that I was coping okay with an intellectual field distant from my own. Dreyfus was screening a film for a class and I went along. That evening, I wrote in my journal:

> *Turned out to be Carl Dreier's Day of Wrath, which I found so riveting I was nearly late to see the people who want to rent our Berkeley apartment. A stunning study in the power of evil, but where does evil lie? In female sexuality? In men's weakness in the face of it? It raised many questions.*

In *Machines Who Think,* I treated Dreyfus respectfully but also told the truth, which often made him look foolish. When the 25th anniversary of the book was to be published, I emailed him, asking permission again to use the quotes from his book that I'd used in the original edition. He insisted I call him. On the phone he told me I decidedly did *not* have permission, and furthermore, now that he was retired, he'd been talking it over with his friends, and was seriously considering suing me for defaming his character.

"That book's twenty-five years old," I said, starting to laugh. "Nevertheless!" he cried, shimmering with such indignation he couldn't finish the sentence. "I'll wait to hear from your lawyer," I said, and hung up before I was convulsed. The new edition went to press without those quotes.

Dreyfus wasn't done. As I was promoting the new edition on a morning call-in show in San Francisco, he was first on the phone.

Gleefully, he told me and the radio audience that the recent DARPA self-driving car competition had ended in a rout; on the 142-mile course, the best car had gone only 7.4 miles. This was proof that machines could never…etc. I explained patiently on the air that science was often incremental, and that maybe next year the best car would go ten miles, and later fifteen, and so on. In fact, the following year, 2005, several cars completed the course, with Sebastian Thrun's self-driving car in the lead. Nowadays, nearly all automobile companies have prototypes (and Dubai has even announced plans for flying "drone taxis that skip drivers and roads" using a Chinese-made vehicle [Goldman, 2017]). Legislators worldwide ponder what the rules of the road should be for autonomous vehicles. But my phone didn't ring with an apology, because Dreyfus could never, ever utter the words, "I was wrong."

In the Winter 2013 issue, *California*, the University of California alumni magazine, ran a brief sidebar about Dreyfus's quixotic fight against AI and quoted a statement he made in 2007: "I figure I won and it's over—they've given up."

Over all these years, I've suspected Dreyfus of many things, but a sense of humor?

Dreyfus died in Berkeley on April 22, 2017.

5.

In the mid-1980s, I met Vartan Gregorian, then the head of the New York Public Library, at an international PEN meeting in New York City. I introduced myself to this distinguished-looking, gray-haired, neatly bearded presence, his deep dark eyes containing all the sorrows of the Armenian diaspora. "I know who you are," he cried with surprising glee. “I’m giving a party for Bert Dreyfus at the Library next week. I’ve read all the literature. I insist you call my office with your address so we can send you an invitation.”

I was stunned. Why a party? “Because I’m a Romantic, and I like Bert’s idea that machines will never be interchangeable with humans.” Is that what he thought I was about? Humans interchangeable with machines? Me reduced to a facile formula that bore no resemblance to what I’d thought or written? I didn’t even think one human being was interchangeable with another. Mark Harris’s dictum rushed to mind: “A writer should not run just for local office.” I didn’t call Gregorian.

John Searle came to speak at a Columbia University convocation in 1981 and presented the Chinese Room Argument. A computer (or the philosopher himself) is isolated in a room with slips of paper, on which are written Chinese characters. He must translate Chinese into English, matching character to word without having the least “understanding” of Chinese. All he does match a symbol in one language to a symbol in another. He produces a translation, but if he doesn’t “understand” what he’s doing, then the act doesn’t qualify as intelligence in any sense.

From the rostrum we strolled along College Walk together, and I told him how disappointed I was that a challenge as substantial as the Chinese Room Argument didn’t exist earlier. But he’d only begun thinking about AI the year after *Machines Who Think* was published.

Philosopher Daniel Dennett at Tufts University and computer scientist Doug Hofstadter at Indiana University made the first plausible attack on the Chinese Room Argument in their book, *The Mind’s I* (1981), and you can read the history of post and riposte over the past decades in Dennett's delightful *Intuition Pumps and Other Tools for Thinking* (2013).

To summarize, the isolated computer, or, for that matter, human philosopher, cannot translate Chinese character for English word, one-to-one, after all. Fundamental to language translation is real-world knowledge, just as it is to most linguistic transactions. However, thanks to the vast amount of data on the Internet, machines can now acquire considerable real-world knowledge, as the program Watson showed when it triumphed over the best human *Jeopardy!* players in 2011. Watson's win required not only real-world knowledge, but also the ability to catch puns, jokes, and other subtle linguistic properties.[2] Did Watson really "understand" what it was doing? Or was the machine only an example of "weak"—albeit pretty dazzling—artificial intelligence (which Searle was okay with)?[3] The Chinese Room Argument was constructed on venerable but misleading philosophical tradition that for intelligent behavior, reasoning was far more important than knowledge.

To make matters worse for the Chinese Room Argument, in October 2012, Rick Rashid, then head of Microsoft Research, gave a lecture in China to demonstrate software that transcribed his spoken English

[2] In January 2013, reports circulated that Watson had been hooked into the Urban Dictionary, a crowd-sourced, online, up-to-the-nanosecond dictionary of slang, used by teenage boys and certain elderly connoisseurs of living language like me. One of the Watson team developers thought that Watson should be more informal, conversational, hip. But when Watson answered a query with "bullshit," team members decided to purge the Urban Dictionary from Watson's memory. I haven't checked this story. I can only hope it's true.

[3] In The Quest for Artificial Intelligence, Nils Nilsson (2010) observes astutely that when Herbert Simon's children enacted the Logic Theorist's moves and proved a theorem, the children's understanding was in doubt and yet the theorem was indeed proved. Philosophers might cry out on behalf of intentionality, but if you're going to ascribe intentionality to every cell, it gets mighty complicated.

words into English text with an error rate of about seven percent. Then the system translated them into Chinese-language text (error rate "not bad," Rashid would tell me in late 2015) and followed that with a simulation of Rashid's own voice uttering them in Mandarin. A real Chinese Room: you can see it on YouTube.[4] It wasn't perfect, but as Rashid said to me later, with all the examples to learn from on the Internet, it's much better now and improves daily.[5]

What do we mean by understanding (yet again)? Only humans can really understand, Searle has argued, because they exhibit "strong," not weak intelligence. Thus Searle says "strong" *artificial* intelligence contradicts itself. Only humans can have strong, or real, intelligence, because only humans understand. Whatever that is. It must be the *wonder tissue* in our heads, says philosopher Daniel Dennett with a wicked grin. (But then, tens of petaflops of processing on 20 watts of energy, as the human brain exhibits, is pretty wonderful.)

As I write, machines have all but closed down the Chinese Room Argument and similar hypothetical problems in text and are whizzes in facial recognition, better than most humans at reading the emotions of other humans, better than any humans in molecular recognition and generation (for molecular biology) and in image

[4] See https://youtu.be/Nu-nlQqFCKg

[5] Rashid told me about this at a symposium to honor the fiftieth anniversary of Carnegie Mellon's computer science department. He also told the symposium's audience that, at that lecture, some members of the Chinese audience had wept with joy to hear such a momentous thing. Google's percent of word error dropped from 23% in 2013 to 8% in 2015. Similar improvements were apparent in image recognition and machine translation from one natural language to another. See Dietterich, Thomas G. (2017, Fall). Steps toward robust artificial intelligence. AI Magazine, 38(3) pp. 3-24. doi: https://doi.org/10.1609/aimag.v38i3.2756. A current challenge is to understand spoken words that mix several languages.When President Donald Trump visited China in 2017 and gave a public speech, the translator was now a Chinese program called iFlytek, a demonstration of how quickly China was climbing the AI achievement ladder.

recognition and generation.[6] They're beginning to read human brain messages and transmute them into physical action, an answer to the question that puzzled early AI researchers: how does intelligent behavior emerge from dumb tissue, or dumb components of any kind? They're capable of many other useful applications, employing what is known as deep, or multilevel, learning. But they're still machines, woefully deficient in wonder tissue.

6.

Back in the early 1980s, my husband Joseph Traub was also provoking people. As the founding head of the new computer science department, he'd been invited to address Columbia College alumni and spoke to a packed hall on the topic of whether computer science was a liberal art. He argued yes and stunned—perhaps insulted—the deep core of humanities professors and former and present-day students, who, not surprisingly, thought of computer science only as Coding 101 and computers themselves as nothing but big, dumb machines (as ads from IBM kept reassuring them). You can sympathize with their disbelief. The Ivy League had been late coming to computer science, and Columbia was one of the last of all.

New York City might have been the cultural capital of the free world, but computationally speaking, I'd taken Joe from an advanced civilization that existed in maybe three places on the planet and brought him to a windowless sod hut on a desolate prairie. To

[6] Image recognition and generation is a razor-edged sword: so useful for so many applications but so good at generating fake images that soon you won't be able to believe your own eyes. The same week "Afterimage," a long article on this topic by Joshua Rothman, appeared in The New Yorker (November 12, 2018), the White House itself was accused of using a doctored video to justify suspending the press credentials of an aggressive reporter. The doctoring was a speeded image, needing no AI, but the implications of the forensics were deeply disturbing.

transform that sod hut, Joe faced a mighty task. For this, I felt deeply sorry. He never reproached me for taking him away from the bright lights of his own field to the bright lights of mine.

But time passes. In February 2014, Columbia University's *The Record* celebrated the university's Digital Storytelling Lab, which brings together statisticians, English professors, filmmakers, and social scientists "to tell stories in unexpected and, sometimes, never-before-imagined ways" ("Humanities cross," 2014). That same issue of *The Record* also profiled Alex Gil, Digital Scholarship Coordinator, Humanities and History Division, a part of the Columbia Libraries. He helps Columbia faculty to use digital technology in humanities scholarship and teaching (Shapiro, 2014).

A second profile in that issue of *The Record* was of Dennis Tenen, assistant professor of English and Comparative Literature, whose brief is digital humanities and whom you met in Chapter 4 when he addressed a group of Harvard scholars and suggested that intelligence might reside in the system as much as in a human heads—and didn't promptly get booted out of the seminar room. Tenen told *The Record* he was at work on a book about algorithmic creativity (think the sonnet form) and was devoted to understanding culture through a computational lens and computation as a cultural experience (Glasberg, 2014).[7] As we'll see later, nearly all major American

[7] Dennis Tenen's 2017 book about algorithmic creativity is Plain Text: The Poetics of Computation (Stanford University Press). Forgive a certain unseemly triumphalism here. For thirty years, one dean in that early Columbia audience of Joe's, considering me the friend of his enemy, mustered the sourest, angriest look he could whenever we encountered each other in Morningside Heights. In 2014, it must have been bitter news to read that the president of Columbia, Lee Bollinger, said in an interview in the Spring 2014 issue of Columbia, the university's alumni magazine: "Ten years ago, our engineering school was at the periphery of the University, and its faculty members, I'm told, felt unappreciated. Now they are at the center of intellectual life on this campus." Bollinger hastened to add that so too were the business, journalism, and public health

universities and many in Europe now have equivalent centers and similar scholars. Professional organizations and journals flourish.

7.

Computer scientists themselves didn't always appreciate how intellectually rich computers would prove to be. Almost thirty-five years would pass after Joe's talk before we'd read anything like what Leslie Valiant, of Harvard's computer science department, writes in *Probably Approximately Correct*:

> Contrary to common perception, computer science has always been more about humans than about machines. The many things that computers can do, such as search the Web, correct our spelling, solve mathematical equations, play chess, or translate from one language to another, all emulate capabilities that humans possess and have some interest in exercising. . . .The variety of applications of computation to domains of human interest is a totally unexpected discovery of the last century. There is no trace of anyone a hundred years ago having anticipated it. It is a truly awesome phenomenon (Valiant, 2014.)

As these examples show, dissenters fell into several categories. Many scientists in distant fields felt moved—threatened?—enough to show why, by their lights, AI couldn't be done. In the case of Arno Penzias and others, the empirical evidence that might contradict their beliefs wasn't even worth examining. Nor did most philosophers respond to empirical evidence: in their hearts they knew it couldn't be done, so constructed parables to prove it. In the case of Vartan Gregorian, he simply misunderstood—Pamela McCorduck, at least, did not think machines and humans were interchangeable—and went with fast

faculties, so I suppose the computer science faculty shouldn't get a collective big head. "But data science is certainly a dominating force of our time, one that is having a transformative effect on many fields" ("The evolving university," 2014, p.31).

thinking, his Romantic impulses. Someone like Joe Weizenbaum believed it could be done, but no amount of good AI might do would compensate for its potential evil.

In July 1999, my husband and I went to Oxford for an international meeting on the foundations of mathematics. Knowing nothing about the topic, I planned to be a carefree Oxford tourist, admiring the greens and Gothic spires. But to my surprise, this exceptionally abstruse meeting featured a panel on "Computation, Complexity Theory, and AI."

I joined Joe in the plenary audience wondering why the panel had no expert in AI. Two exalted mathematicians sat onstage: Richard Brent, an eminent theoretician in complexity who'd stepped in for Tony Hoare, who'd mixed up the date; and Stephen Smale, a Fields Medalist and specialist in some of the more arcane parts of mathematics. With them was one physicist, Roger Penrose. However, Penrose had recently published a second book saying why, for reasons of quantum physics, AI was hopeless.

I'd read the first of Penrose's books attacking AI—or tried to. The parts about quantum physics seemed right, at least as far as I could judge, but the parts about AI seemed shockingly ignorant. Maybe, I thought, he knows something about the other topics the panel means to address, computation and complexity theory.

From my journal, July 27, 1999:

> *It's the usual physicist-twit's view that he can come in and clean up the problems in any field whatsoever, but alas, knowledge counts, and Roger P. knows zilch about any of this. As Richard Brent says privately later, it's as if his knowledge of all three topics began and ended with Alan Turing. Richard also suspects Penrose has religious reasons for his antipathy to AI,*

but this we don't know. On the whole, Penrose is a slightly more interesting adversary than Bert Dreyfus, but not more convincing. In fact, less. For he keeps proposing experiments that "can't yet be done but in the future..." or "I'm assured could be done..." or "experiments might be performed..." etc., all this to prove/disprove what he calls "my position," which turns out to be the most ridiculous sort of phrase-dropping and general obfuscation. Big emphasis on "consciousness" as essential to intelligence, by which I take him to mean self-consciousness. He's so innocent of intellectual history that he doesn't realize "consciousness" in the sense he means is a cultural construction, missing in great parts of the human world to this day (so I guess they aren't "intelligent") and only making its first appearance in Renaissance Europe. Blech.

So WTF is this all about? I think of standing up and informing the audience that I've never been to an AI meeting (and I've been to plenty) where panels were convened on "Partial Differential Equations: Fact or Fiction?" or "Why Don't These People Understand That Reynolds Numbers Don't Help Navier-Stokes Calculations of Turbulence?" The whole performance is bizarre, another example of AI-envy disguised as sermon, cold shower, neener-neener. The house is packed, of course. My old argument that rhetoric is beside the point in science occurs to me, but then I think of the Lighthill Report, that more or less killed AI funding (hence research) in Britain, and can possibly be held responsible for the dismal state of computing here. The Brits were there first, and now they're simply not players. What a price people in the UK paid for that piece of rhetoric. Not that I blame Sir James, especially. He was the author of the blunt instrument, but many hands wielded it, and many more refused to rise up and stay that blunt instrument, all complicit.

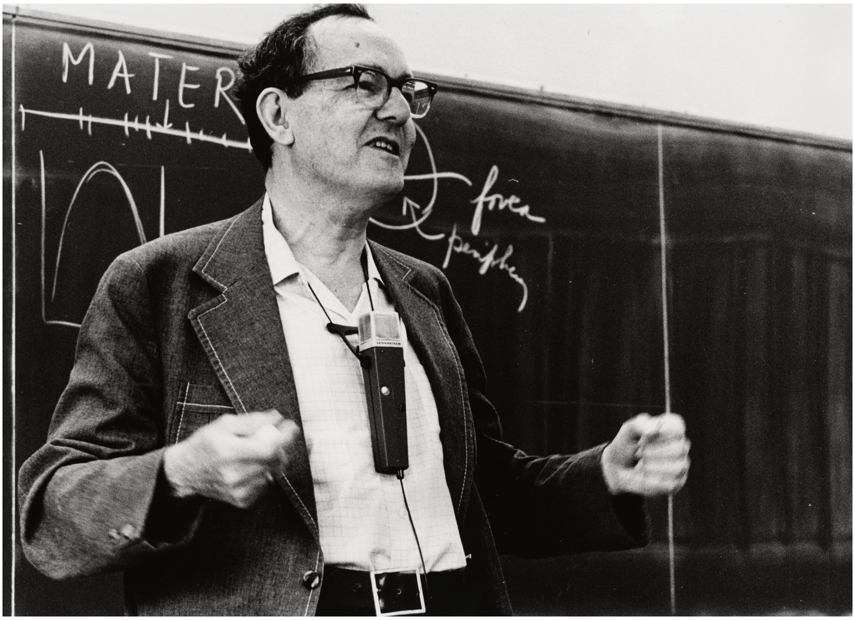
Joe actually took on Penrose for his misstatements about complexity—computational complexity, since he's obviously completely ignorant of other forms of the genre—but I thought it wasted breath; this man is not open to contradiction or even learning. So much for "intelligence."

> *The upshot of the panel was that AI is, as usual, barking up the wrong tree, premature, blah blah. I could only laugh.*

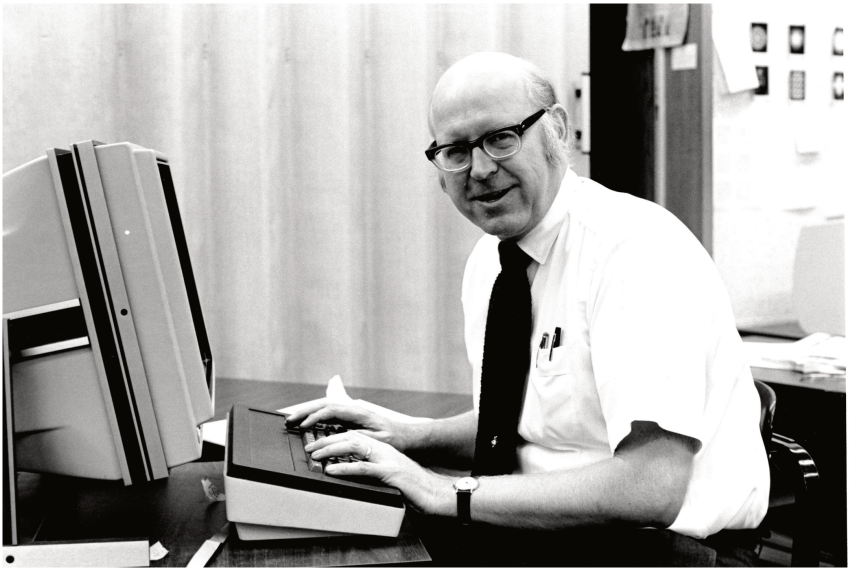
As a former subject of the U.K., and sentimentally attached, it gives me pleasure to report now that the London firm of DeepMind is in the avant garde of AI. Twenty years ago, it wouldn't have seemed plausible.

IBM's Watson beats past Jeopardy! winners Brad Rutter and Ken Jennings in 2011.

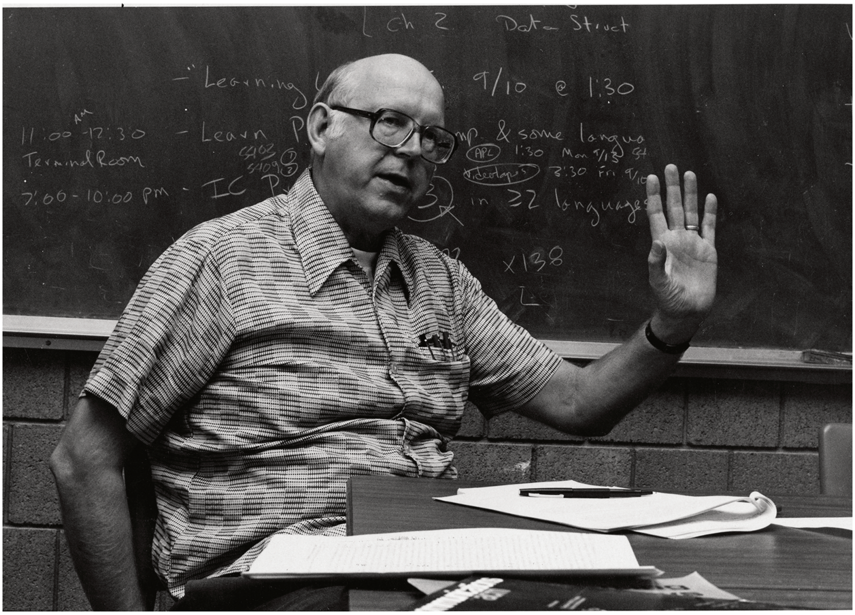
Herb Simon plays chess with CMU faculty member Bill Chase in 1973. Neil Charness, a PhD student who worked with Simon on his chess experiments, films.



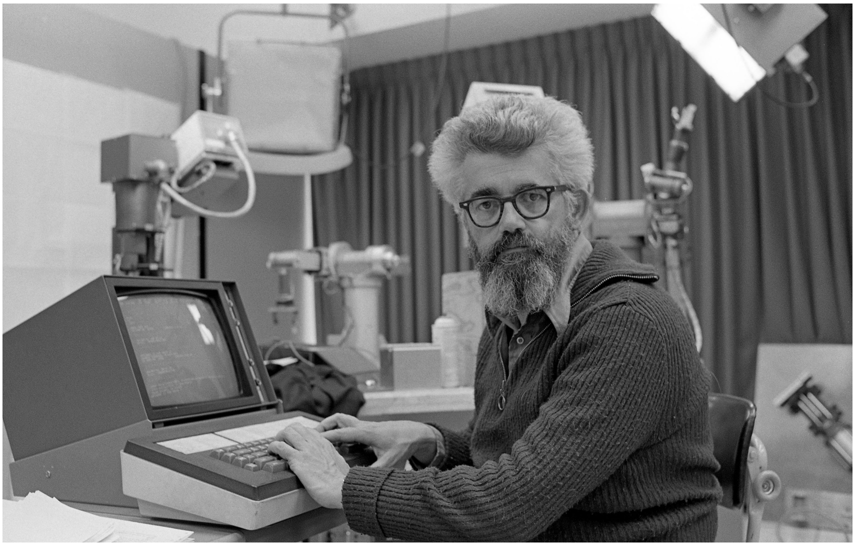
Herb Simon lectures in Hamburg, Germany in June 1977.



Allen Newell sitting at a prototype computer with a CRT monitor, designed to provide visual feedback to the user, 1975.



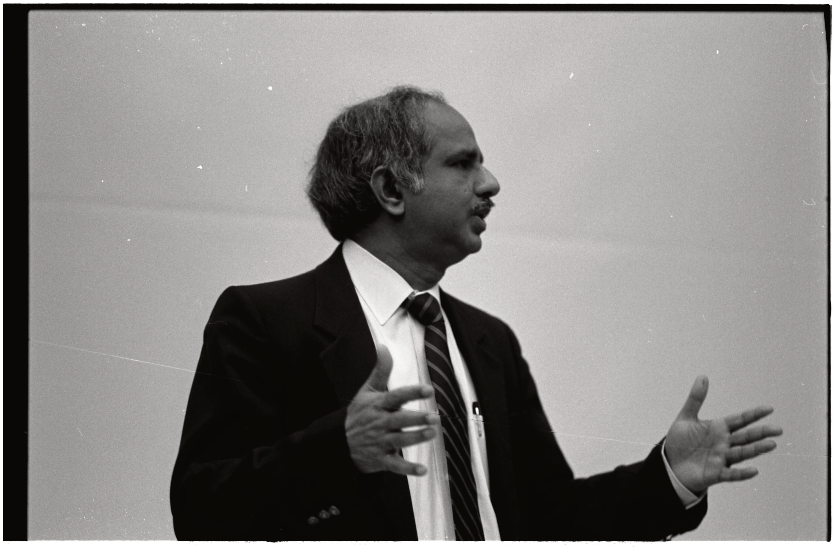
Allen Newell teaches a seminar course in 1977.



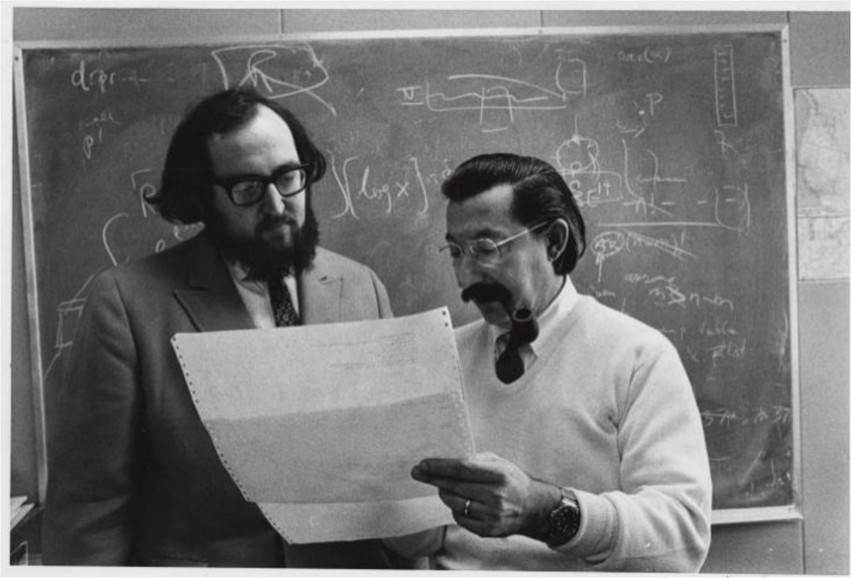
John McCarthy, one of the founding fathers of Artificial Intelligence.

Marvin Minsky and the “Minsky Arm” at the MIT Artificial Intelligence Lab.

Edward Feigenbaum (center), Director of the Stanford University Computation Center in 1996 with Gio Wiederhold, Bob Braden, and an executive at the Stanford Computing Center.



Raj Reddy meets with students at Carnegie Mellon's Graduate School of Industrial Administration (GSIA) in 1989.

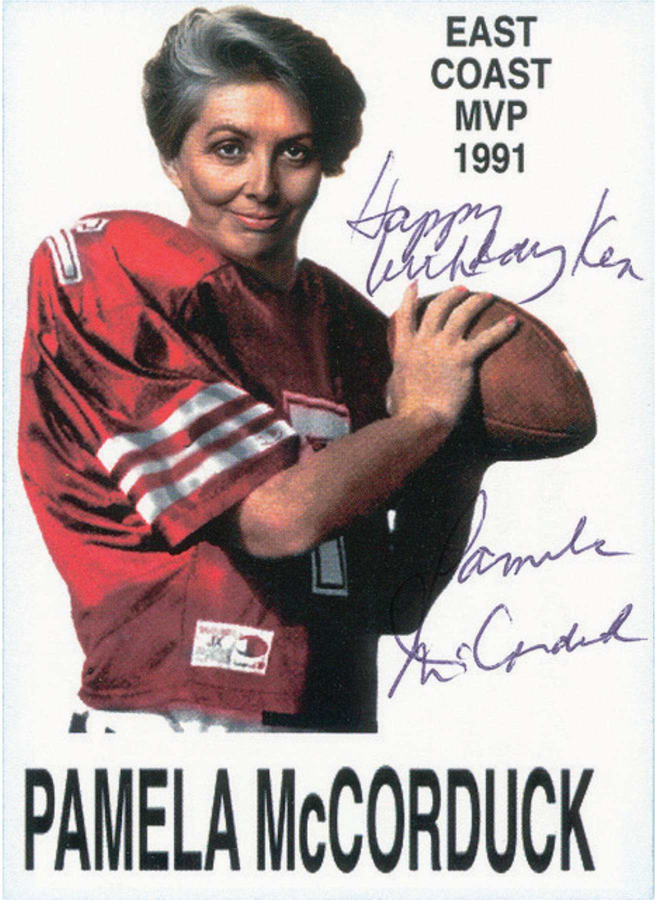


Joel Moses (L) and Joseph Weizenbaum (R)

Maja Mataric

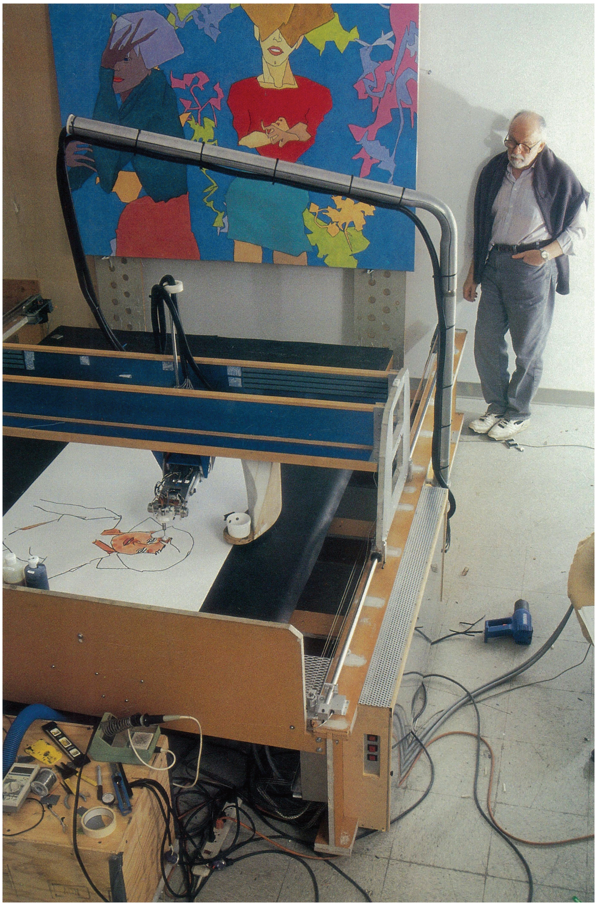
Pamela McCorduck and Ashley Montagu at the opening of Santa Clara University's Technology Institute in January 1986.

Lofti Zadeh



Pamela McCorduck's 1991 East Coast MVP trading card from the Computer Museum of Boston's yearly Computer Bowl competition.

Apple CEO Steven P. Jobs, left and President John Sculley present the new Macintosh Desktop Computer in January 1984 at a shareholder meeting in Cupertino, California, USA. (AP Photo)

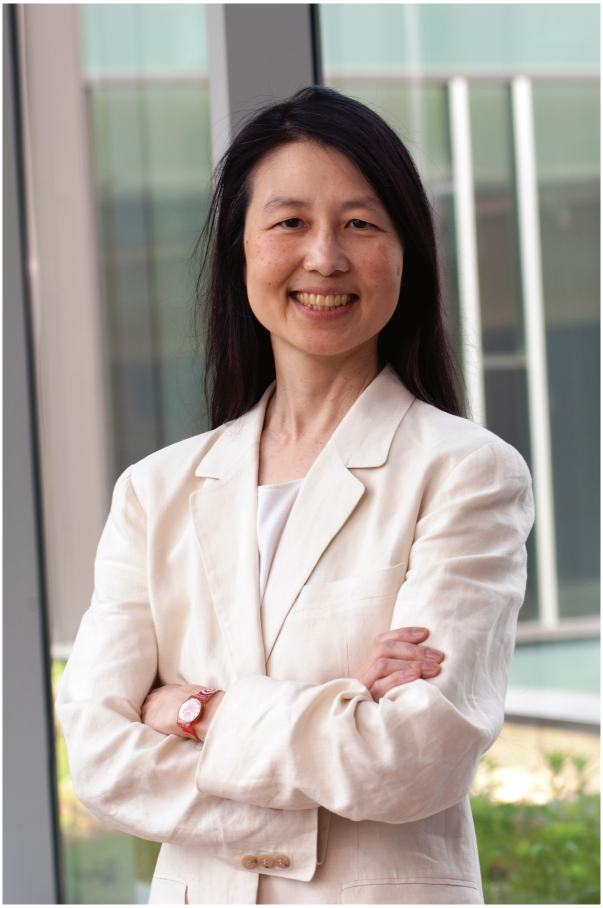


Harold Cohen in his studio.

Patrick Winston at desk

Elizabeth Honig addresses a conference in 2014 about her work .

Daniel Dennett in the Fall.



Jeanette Wing, 2010

Mary Shaw

Kai-Fu Lee

Part Four: The World Discovers Artificial Intelligence

As though
the river were
a floor, we position
our table and chairs
upon it, eat, and
have conversation.
As it moves along,
we notice—as
calmly as though
dining room paintings
were being replaced—
the changing scenes
along the shore. We
do know, we do
know this is the
Niagara River, but
it is hard to remember
what that means.

—Kay Ryan, "The Niagara River"

Japan Wakes the World Up to AI

1.

Sometime in Spring 1982, Ed Feigenbaum and I were having one of our phone chats, me lamenting that a book on computer graphics I'd begun was about to be pulled out from under me by the publisher who proposed it because his underwriter was going bankrupt.

Ed said, you know? The Japanese are doing something really interesting. They're putting big resources into the next generation of AI. It really might vault them ahead.

Ed sent me a few documents describing what the Japanese hoped to do, and he was right. This was big news. The Japanese Fifth Generation group had decided the time was right to move ahead significantly in AI, design an epoch-making computer to run it, and now had the backing of mighty MITI, the government Ministry of International Trade and Industry. I wrote a proposal for my new agent and went to Utah on a ski vacation.

The Fifth Generation would turn out to be a grand adventure and awaken much of the world to AI. Even the First Culture would finally deign to turn its head and notice, with mixed results.

The moment was good for AI, which was getting more and more public notice. The moment was also good for anything about Japan, because in the early 1980s, a number of books emerged to claim how much smarter the Japanese were in manufacturing and trade than Americans. Ed and I were going to tell the world that their computers might be smarter too.

The Japanese had based their next generation system on Feigenbaum's recently discovered and empirically proved knowledge principle: intelligent behavior arises from specific and deep knowledge, not just reasoning power, as discussed in Chapter 13.

My new agent, John Brockman, called me at a ski lodge to tell me the proposal had already gone out to sixteen publishers and an auction for the right to publish it would take place on March 31. "Are you skiing?" he asked. "No," I said, "I'm inside reading *Bleak House*." "Good. Keep safe." I thanked him, went back to my breakfast, and tried to gather my wits.

When I got back to New York, I smiled to see that several of the publishers who planned to participate in the auction had turned down *Machines Who Think*. Times had changed. Addison-Wesley won the auction and was enthusiastic, urging us on. That spring and early summer I worked feverishly. On my small electric typewriter—I'd buy my first computer with the royalties from *The Fifth Generation*—I hammered out a first draft. Ed and I collaborated in a way congenial to us both, him with some big ideas (and much first-hand knowledge of Japan, its computer industry, and its education system); me with the questions, awkward and otherwise; the beady eye; the skepticism; a sense of larger context and connections; fascination with the people involved—and willing to write. At the end of July, I made my first visit to Japan, a country I'd love evermore.

2.

In Tokyo I rendezvoused with Ed and his Japanese-born wife, Penny Nii, who had received her masters' degree in computer science at Stanford and was Ed's intellectual as well as life partner. She was a knowledge engineer, a vital part of the team to develop any expert system then. A knowledge engineer extracted the knowledge from the head of a human expert and recast it into an executable computer program called an expert system. This task is now automated over large data sets. (Ed claims that Herb Simon was the first knowledge engineer, a man who extracted all the chess expertise from Adriaan de Groot's book on chess masters and recast it as a chess-playing program.)

The headquarters of the Japanese project, called ICOT (Institute for New Generation Computer Technology), was in a generic high-rise with inspiring views of Mount Fuji in the distance. Japan's government had dragooned eight of the leading Japanese electronics companies to participate in the project, each contributing researchers. Not all firms participated willingly. The project leader, Kazuhiro Fuchi, wouldn't allow the reluctant companies to fob off second-rate researchers; he had final say on who'd work at ICOT. He was quite un-Japanese, the strength of will emanating from him like a force field. Although he received Ed, Penny and me in his nicely furnished office with the snow-capped Fujiyama out the window, I knew from one of his researchers, Toshi Kurokawa (he and his wife, Yoko, had earlier translated *Machines Who Think* into Japanese) that Fuchi usually worked in a crude little cubicle where he could oversee his troops. Despite Fuchi's formidable will, the eager talent of these young scientists, and the backing of MITI, the whole thing seemed to me terribly fragile.

Over the next few days, Ed, Penny and I visited participating firms. At Hitachi, we heard that they hadn't wanted to join in this wild scheme; MITI had "made them do it," whatever that meant. They saw themselves as "followers of IBM," and only when IBM felt AI was worth doing would they also willingly do it. At the other end of the spectrum was NEC, determined to do everything to make the Fifth Generation succeed.

Ed and Penny were demigods in Japan, and I was considered a kind of retainer in their wake. No wonder. Expert systems, with their real-world applications, were sweeping AI in Japan as well as the United States (and elsewhere), so a visit from Feigenbaum and Nii was a descent to earth of beings who normally dwelled in celestial regions.

Ed and Penny repaid this rapturous devotion by offering infinite patience with their disciples (they had many in every company we visited). After each talk, each demo, they asked careful questions, gave guidance, and abundant encouragement. The general ideas of expert systems weren't too difficult to grasp; you could attack problems at many levels of expertise and at the end have something to show for it. Thus the most gifted people in AI did the innovative work but left opportunities for the less brilliant to make useful contributions.

However, this method only worked when the deities came around and did regular evangelism, patting, encouraging, and applauding. "Yeah," Ed said wearily late one night. "Sometimes I feel like the slab in *2001: A Space Odyssey*. I come down from time to time to see how everybody's doing. I see they're not there yet, but I tell them to keep at it and go away until the next time."

On one of our many long car trips from an outlying firm back to Tokyo, Ed told me that it had taken much missionary work to make people see the value of actual knowledge to make thinking machines successful. "I took my cue from Herb Simon," he said. "Anytime Herb had an opportunity, he'd write a paper, or give a talk, popularize his ideas, and put them into language that a given audience could understand, tell them why it was significant to them. It was—it is—exhausting work, but the only way to have an impact."

3.

The Fifth Generation was fun to write. Ed and Penny knew much about Japan, and I was learning. Within eighteen months, I brought the finished manuscript to my agent John Brockman at a local falafel stand—John was never big on fancy literary lunches—carrying both copies in a couple of Zabar's shopping bags.

The editorial back-and-forth was more daunting than usual, because our assigned editor was determined to snuff from the manuscript every possible sign of life. V. S. Pritchett had long ago told me that I'd been lucky to publish my first book in England. English editors welcomed the writer's idiosyncratic voice and tried only to make sure the prose was reasonably clear. American editors were "ridiculously meddlesome."

The whole process was somewhere between melodrama and *opera buffa*. The publishers had bought our names and ideas but wanted to write their own book. The editor was ungifted and oppressive. ("Is editorial heaven," I asked this man sweetly, "a place where manuscripts appear without the inconvenience of authors?") The revisions he insisted on were so awful that from a Caribbean holiday, Ed emailed me: "I've dragged this thing around like a dead dog: I can't read it, and I can't get rid of it." Matters got so bad that Ed and I finally told the editor-in-chief that we were withdrawing the book and would of course return the handsome advance.

This got editorial attention in a useful way. The oppressive editor was fired ("I was leaving anyway") and I attempted restoration. Writing with original verve is one thing, but I soon realized why resurrection is properly considered a miracle. Even the final product appeared with a blunder so big on the printed cover that we insisted the publisher recall the book and fix it. Eventually, in authorized and unauthorized editions, in many languages around the world, the book would sell very well.

4.

In the long run, the Fifth Generation didn't quite turn out as the Japanese hoped. Some argue it was before its time (its multiple levels of programming anticipated deep learning). Others say that the evolution of off-the-shelf components made obsolete the special-purpose machines the Japanese proposed to build or that the programming language chosen was too cumbersome.

In the only English-language evaluation of the Fifth Generation, Ed Feigenbaum and Howard Shrobe (1993) of MIT laid out the project's detailed technical achievements and failures. To wit, the project did little to advance the state of knowledge-based systems, AI as such. Its natural language goals and other human interface goals were dropped, and its hopes of useful applications did not materialize. The Fifth Generation project's research and development of parallel reasoning machines, as opposed to linear machines, was almost unique and very useful to parts of AI that require heavy signal processing (vision, speech, robotics). However, lack of parallelism wasn't the biggest challenge for AI; the lack of ways to deal with large-scale knowledge bases was a much bigger problem. (This was to change, but not for another decade or so.)

But ICOT's achievements showed that Japan could innovate in computer architecture; at peak performance, its specialized machines reached the original goals the project set. Above all, the project created an attractive aura for AI, knowledge-based systems, and innovative computer architecture. "Some of the best young researchers have entered these fields because of the existence of ICOT," Japanese scientists reported. Japanese roboticists are now world leaders.

And yet. After the deadly Tohoku tsunami, when the Fukushima Daiichi atomic power plant melted down in 2011, Japanese robots should have been on the spot. Unfortunately, government agencies that guided and funded research had believed a disaster on the scale of Three Mile Island or Chernobyl could never happen in Japan; thus government decision-makers saw no point in financing research into robots that could withstand high levels of radiation. It was a tragic miscalculation with ghastly consequences.

Without counting the human cost (as if you could), Fukushima Daiichi is a giant demolition project that will require an estimated 40 years to complete and cost $15 billion. Three years after the meltdown, demolition work began. Robots that could climb over debris were deployed but had to be controlled by cables that got easily tangled because radioactivity interferes with wireless transmission. Planned next were robots that could cut through obstacles and pick up debris, followed by janitor robots that used high-pressure water jets and dry ice to clean wall and floor surfaces (Strickland, 2014).

The spent fuel rods must be removed, the radioactive water contained, and finally, the three damaged nuclear cores removed. That job alone might take twenty years. For the most part, humans control the robots I've described, and improved autonomous robots might be more successful—but nothing can speed the half-life of the radioactive cores. In 2017, a small (shoe-box sized) radiation-hardened robot was built and deployed to find the melted down fuel cores. Like an aerial drone, Manbo used tiny propellers to navigate through the radioactive water used to cool down the reactors and finally found and videoed the three reactors whose cores had melted down during the disaster (Fackler, 2017).

Alarmed by the failure of conventional robots at the time of the disaster, the U.S. Department of Defense's Advanced Research Projects Agency (DARPA) conducted a grand competition in robotics between 2013 and 2015, aimed at producing radiation-proof autonomous robots to go into catastrophic sites like Fukushima, open doors, move debris out of their path, turn off valves, climb ladders, connect a firehose to a hydrant, and perform other difficult tasks in a human environment. The competition kept roboticists up late all over the world, subsisting on Cokes, pizzas, and instant ramen, working to win. Roboticists at SCHAFT, a firm owned by Google but originally Japanese, handily won the midterm round of challenges in December 2013, but in June 2015, a team from South Korea's Advanced Institute of Science and Technology won the grand competition (and its prize of $2 million) with a humanoid robot, DRC-HUBO. In 44 minutes and 28 seconds, the robot completed all eight of the competition's tasks flawlessly (Guizzo & Ackerman, 2015).

As we know, catastrophe can happen anywhere.

5.

One reason, though hardly the main one, the Japanese were keen to ramp up AI in the early 1980s was a stark demographic fact: the Japanese population was growing older rapidly, and this large cohort of elders must be cared for. Perhaps because of this, certainly because I thought our book was getting tech-heavy and needed some levity, I introduced the Geriatric Robot.

The great thing about the Geriatric Robot, I wrote, is that it doesn't just clean you up, feed you, or wheel you out into the fresh air. The great thing is that it *listens*. Tell me again, it says, about how wonderful/awful your kids are. Tell me again, it says, about that great coup of '73. It listens patiently and sincerely—again and again. It isn't hanging around to inherit your money or because it can't get any other job. You *are* its job. It doesn't get distracted or bored. It doesn't judge you. It's an attentive caregiver who will be there long after your biological family has lost its serenity or your hired help is fed up. "We humans can't help it," I added. "It's part of our charm" (Feigenbaum & McCorduck, 1983).

In the past few years, I'd often been invited to give talks to college students and needed to illustrate AI with something that would be vivid to people that age, and better, make them laugh. My dear friend, the novelist Hortense Calisher, who was in her seventies, thought the Geriatric Robot was hilarious and ought to find a wider audience. If Hortense, at her age, didn't find it offensive, then I imagined other people wouldn't either. In the book, I flagged it with all sorts of rhetorical signals that I was just kidding.

Ed took a look at it and said, maybe not. I insisted. The editor excised it. But that editor had tried to throttle every sign of life the manuscript showed, so I put it back. Fun or not, it seemed appropriate, given the Japanese plans to meet responsibilities of eldercare with AI.

The Geriatric Robot was a small part of the book—nothing compared to other, more significant challenges raised by the Japanese, which soon brought an invitation to Feigenbaum to testify before a Congressional hearing—but for the First Culture, it was proof positive that AI people, me included, occupied in the Great Chain of Being the level of insensible brutes.

That's just above the plants.

Stragglers from the Wreck of Time

1.

A month or two after *The Fifth Generation* was published, Mike Dertouzos, an intense, Greek-born man of ebullient cheer, then the director of MIT's Laboratory for Computer Science, came to Columbia for some event. We sat next to each during lunch at The Terrace restaurant, gazing from Morningside Heights over Morningside Park, Harlem rooftops, and the East River. "Beware," Dertouzos said with uncharacteristic seriousness. "Weizenbaum has apparently reviewed *The Fifth Generation*, and he's walking around the halls, rubbing his hands together, telling anyone who'll listen that *this'll get 'em*."

Joe Weizenbaum had originally engaged in a mid-1960s fight with Kenneth Colby about the use in mental hospitals of Colby's program, Doctor. Colby was trying to bring some automated relief to patients who, if they were lucky, saw a psychiatrist once a month. Weizenbaum asserted that no machine must interfere in the psychoanalytic process, and if that meant patients went without any therapy, so be it. This eventually led him to attack the entire field of AI, steadily and publicly.

Unlike philosophers of the time, who maintained AI could never succeed, Weizenbaum said to the contrary, the effort certainly could succeed. His early work on Eliza, the therapist program, and his affiliation with MIT gave that opinion public gravitas. But he believed strongly that it mustn't be done and was the fast chute to catastrophe. To make his points, he leaned heavily on comparisons between AI and the Holocaust.

In 1976 he'd put these arguments together in his book, *Computer Power and Human Reason,* described in Chapter 17 of this book. *Computer Power and Human Reason* was influential and warmly welcomed by people already uneasy at the whole idea of computers, never mind machine intelligence. As a consequence, in the last few years, he'd risen to singular prominence as the self-declared "conscience" of AI and computing in general.

"He used to believe about ten percent of the stuff he was spouting," Dertouzos said over that Morningside Heights lunch, "but now he's become a complete convert to his own line. He's also reading passages aloud to people from the chapter, 'Intellectuals in the Cherry Orchard,' saying how shocking it all is."

"Intellectuals in the Cherry Orchard" was a chapter in *The Fifth Generation* that used Chekhov's Madame Ranevsky, tragically oblivious of the future and her responsibility to it, to make the same point about some present-day intellectuals. But really, I was restating one of C. P. Snow's points, in his Two Cultures lecture years ago. Here's my passage from that offending chapter:

> In short, no plausible claim to intellectuality can possibly be made in the near future without an intimate dependence upon this new instrument. Those intellectuals who persist in their indifference, not to say snobbery,

> will find themselves stranded in a quaint museum of the intellect, forced to live petulantly, and rather irrelevantly, on the charity of those who understand the real dimensions of the revolution and can deal with the new world it will bring about. (Feigenbaum & McCorduck, 1983)

"Weizenbaum's had a hard time since he's come to MIT," Dertouzos continued. "You know—no PhD. So that means you have to be twice as good. As it happens, he hasn't been twice as good. His research has gone nowhere. He knows he's not making the grade. It's got to be very painful. Still, I wish he hadn't chosen this way to compensate."

"Maybe," I said. "But it's also about World War II, the Holocaust. He's possessed by it. Interesting, because he didn't actually go through it himself." He'd told me so during his long phone call from Berlin. I went on to speculate: "So, is that the problem? It's some kind of obsession because he didn't personally experience it?"

Dertouzos and I exchanged our own World War II stories. He'd been a child during the German occupation in Greece. His father was a partisan, and Dertouzos knew from infancy that in that savage occupation, if he betrayed anything that happened at home, he could cause his father's and uncles' deaths and his own. I'd been born in England at the height of the Blitz, into a world where the bombs rained down nightly on the just and unjust alike. So we'd both actually lived through it and more or less put it behind us.

We shrugged. Who knew what was driving Joe Weizenbaum? He'd recently appeared on television, contradicting himself by saying machines could never think because they didn't feel cold and lonely in an empty house. Thirty years earlier, someone had protested to Alan Turing about thinking machines: Impossible! They can never love strawberries and cream! Despite Marvin Minsky's defense of emotion as an integral part of intelligence, would being afraid in a dark house or loving strawberries and cream get you admitted to MIT?[1]

2.

I was warned. I alerted Ed Feigenbaum, and we waited. Sure enough, *The New York Review of Books* soon published a five-page review of *The Fifth Generation* by Weizenbaum (1983). In the opening paragraph, we were compared with Mussolini, Hitler, and Pinochet, my only consolation being that at least Pinochet was still alive. "We wrote a book," I said aloud to the broadsheet in my hands. "We didn't jail, torture, or kill anyone; overthrow any governments. Your editor let you get away with this?"

All in all, it was a review that served neither Weizenbaum nor our readers, although I suppose he got a weight off his chest.

Weizenbaum might have had some technological quarrels with the Japanese—we'd had our own and said so. But his biggest outrage was against me, because what he found most shocking about *The Fifth Generation* was my proposal that AI get busy and do something useful, like the Geriatric Robot. Only people must look after people, he thundered from the pages of *The New York Review of Books*, an echo of his warnings twenty years earlier, that only psychiatrists must conduct the psychotherapeutic interview, even if it meant that souls in torment went without any treatment whatsoever.

"I shouldn't have let you leave it in," Ed said. But he laughed. Soon, at a Manhattan book party for a friend, a stranger made small talk with

[1] Emotions are part of intelligence, Minsky and many others have argued, necessary but not sufficient for intelligent behavior.

me about the review, and when I cheerfully told him it was my book, he backed away as if plague cankers had erupted all over my face.

I wondered if Weizenbaum had ever actually looked after an elderly person or whether he just thought somebody—some woman, let's face it—should. Regardless, demographics suggested that caregivers for the old and infirm looked to be like telephone operators in the 1950s: if trajectories held, half the population would soon be thus occupied. Luckily the Bell System invented automatic switching, and we were all released from compulsory careers as telephone operators.

So while I'd meant the Geriatric Robot as some jokey relief, in fact the telephone operator problem loomed. As a population, as an economy, we simply cannot devote that kind of one-on-one human care to the old and chronically ill. We're all "stragglers from the wreck of time," as Henry James almost put it, or we will be. According to the U.S. Center for Disease Control, in 2018, 61 million Americans had some form of disability. About 65% of people of working age with disabilities are unemployed. By 2030, twenty percent of the population will be elderly and one in two working adults will be an informal caregiver. An acute shortage of trained health professionals in geriatrics will become worse. The Japanese were straightforward about it; why did we have to hide behind such pious humbug?

Ed and I sent a snappy reply to *The New York Review of Books*, which printed it, and that was that. But it wasn't.

3.

I was soon invited to a Japanese conference on robotic help for the aged. I declined with thanks. Then *Der Stern* called from Germany.

Who was doing the software for this? The hardware? A little joke, I said to the stern lady from *Der Stern*. Nobody was doing it, unless she wanted to call Tokyo and follow up further.

But necessity is the mother of invention. Twenty or so years later, Nursebot Pearl appeared, the brainchild of Sebastian Thrun, a young German roboticist then at Carnegie Mellon. Thrun had never heard of the Geriatric Robot, but he deeply loved his grandmother and was sure that, with a little help, she ought to be able to spend her last years in her own home. He had the skills to help her and all other grandmothers threatened with leaving familiar homes of a lifetime.

Nursebot Pearl was experimentally deployed in several nursing homes and centers for the elderly in Pittsburgh, in Cleveland, and elsewhere. She was perhaps four feet high, rolled around on casters like an acting walker, had a sweet and expressive female-ish face, and reminded her elderly clients to take their meds, turn on the TV for their favorite programs, and grasp Pearl's handles so that vital signs could be instantly dispatched to humans on the watch for anomalies. But Pearl became outdated.

In the mid-2000s, an internationally known hardware specialist at Carnegie Mellon, Daniel P. Siewiorek, together with his wife, saw their two sets of parents through a sad and trying end-of-life period. "We can do better than this," he said to himself. CMU received a highly unusual ten-year grant from the National Science Foundation to establish the Quality of Life Technology Center on the campus, pushing a grand suite of technologies to work not just for the elderly, but for anybody, adult or child, with disabilities.

Thanks to partnerships with major corporations, and CMU's strong policy of encouraging such enterprises, some of the Center's products are approaching commercialization. They include robots as home help (one is called Herb, an ungainly looking but useful personal assistive robot for the home) and robots that perceive human emotions and react to them. Virtual coaches have been designed for a variety of disorders, including cognitive and memory assistance, Siewiorek told me. What has been learned in robot vision inspires visual perception enhancements for everyone from the legally blind to Alzheimer's patients to the general aging population—for example, night vision enhancement for safe night driving. Smart wheelchairs help prevent occupants from tipping over, and automobile interiors are being redesigned for the disabled, hand in hand with the designers of self-driving cars.

At least as important is support for caregivers. For example, a relatively cheap and gentle robot with acute sensors might lift and turn a patient over in bed, saving the backs of human caregivers. The design emphasis is on low-cost: apps for smartphones and tablets instead of expensive special-purpose gadgets.

At the University of Southern California, Maja J. Matarić, a professor of computer science, neuroscience, and pediatrics, and her team have been working on robots that take advantage of the wired-in human responses to speech, facial expressions, gesture, movements, and other bio-mimetic behaviors to offer help. That is, these robots monitor, encourage, and sustain all sorts of activities for their clients. They're intended to improve learning, training, performance, and the general health of anyone at risk, whether because of age, autism, stroke, or brain injury. Even the healthy elderly have participated with robotic exercise coaches to keep fit in as pleasant and personal a way as possible.

In the autumn of 2013, Mataric gave a riveting lecture at Harvard, which I attended. She told her listeners she believes that humans respond more deeply to this "embodied" presence, these robots, compared to instructions or conversation on a screen. We're wired, it seems, to assign agency to such a being, so long as its behavior is familiar, interactions with it are believable (not necessarily realistic), and the robot is autonomous (not a guided puppet). Curiously, people who need help respond better to not-quite-perfect robots than they would to perfectionism. "I can't do that," says a vaguely humanoid-looking robot to a human client in one of Matarić's videos, and the client looks relieved. "I can't either."

Matarić shows videos of stroke victims going through their exercises, led by various robots. In stroke rehabilitation especially, motivation is the biggest problem: recovering and using a stricken limb is hard, frustrating, and boring for patient and therapist alike. Even primitive robots of ten years earlier, nothing more than rolling bits of machinery, evoked a relationship with their human clients. Later, more sophisticated robots push, prod, and know, by means of various sensors, when a client is about to quit from frustration. They quickly praise the client's work, and suggest it's time for a break.

Matarić emphasizes the strong effects of the human voice, a problem in AI that hasn't yet been solved to the degree that humans and robots can converse as humans do with each other. Accurate speech understanding isn't enough. Robots will need to enact about seventy nonverbal behaviors too, some as subtle as proxemics—who stands where, and how close?—or the dynamical allocation of roles—now you lead the conversation, now I do.

The uncanny valley effect, first described by Ernst Jentsch in 1906, and elaborated by Freud, was named and quantified in 1970 by Masahiro Mori, a roboticist. Our reactions to something only vaguely human are primarily positive. But when that something reaches a point where it's *almost*, but not entirely human, our comfort level drops dramatically: our positive feelings turn to strong revulsion. Then, as a robot's appearance and behavior cross yet another line and come even closer to human behavior, we ascend from the revulsion of the uncanny valley, and react positively again. The uncanny valley is provoked not just by robots, but by disfigured burn victims, some plastic surgery patients, neurological victims, and even 3-D animations. And, provoked across several dimensions. Matarić says that an Alzheimer's patient grew very disturbed as her robot began to sing like Frank Sinatra. "That isn't Frankie!" she said crossly.

4.

What struck me with Matarić's robots is that they didn't need much intelligence to be valuable in rehabilitation, nor with autistic children, although Matarić is careful to say that not all autistic children respond to robots. But for those who do, the robot is safe and, with its flaws, "like them." Semi-intelligent robots interacting with autistic children can elicit social behaviors, communication, turn-taking, initiating play, and even the first social smile.

So far, economics has prevented scaling up various institutional robots so that we could have one in every older person's home. But as I write, the European Union, facing demographics similar to Japan's, has developed an ensemble of smart clothing, smart environment, plus a personal robot, all to allow the elderly to stay in their own homes longer. The EU plans pilot studies, and also hopes that economies of scale will shrink to a more affordable sum the ten thousand euros that the smart environment and robot now cost.

Canada and, of course, Japan, are working on such systems and robots. Across the United States, centers from Stanford to MIT are studying ways that technology can help. Startups around the country want to remake home care, and adapt social networking for the aged.

I imagine something gradual for most of us—a clever combination of miniaturized off-the-shelf components that deliver smart programs. We'll wear our future geriatric robot as garments or special-purpose prosthetics, wired into our laptops, our smartphones, our baseboards. For all I know, such robots could be implanted. "Apps, gadgets and dongles," says Joe Flower, a specialist on the future of health care, and I'll bet for eldercare, too. Robots around the house could do the boring stuff. Brett the robot, under development at the University of California, Berkeley, named for "Berkeley robot for the elimination of tedious tasks," will learn by apprenticing—watching you or YouTube—and might eventually be available to fold laundry or anything else humans can physically manage.

As I write this, I'm in my seventies, and for all the idealization of human helpers, I've watched some of my friends struggle with the present disadvantages of human home help—while some helpers are magnificent and dedicated, others are poorly trained, paid barely more than minimum wage by their agencies, but cost at least twice that sum out of my friends' pockets. They're on their cellphones disruptively or want the TV on nonstop. I say the Geriatric Robot, whatever its form, will be just in time. Sign me up.

5.

Joe Weizenbaum and I were still not finished. In 1984, I was invited to give a talk at the annual summer meeting of Ars Electronica in Linz, Austria. After an arduous trip, I fell gratefully on my bed in instant and deep sleep. Two hours later, the phone rang. It was Weizenbaum, also at the meeting, and he wanted to talk to me. We agreed on Linzer torte and coffee in an hour.

By now, Weizenbaum was a celebrated fixture on the lecture circuit, warning how the foolish use (which seemed to be any use whatsoever) of computers would inevitably lead to another Holocaust. He regularly preached that for moral and ethical reasons, people must be persuaded to abjure computers altogether.

In the passing years, his hair had grown longer and grayer, his eyebrows more bosky, his moustaches more salient, the pouches beneath his eyes more bulging. Maybe because he was spending much time in Berlin and speaking German regularly, his German accent in English was even more Teutonic. He was the very model of the modern prophet of doom.

Over the Linzer torte, I sat across from him and waited. He made desultory small talk—how was my trip? I waited some more. I began to sense that he wanted to make peace, but he couldn't bring himself to apologize for, or even explain, his disproportionate attack on *The Fifth Generation* in *The New York Review of Books*. Where was peacemaking to begin? He said something about how tragic the Holocaust had been, and if we didn't watch out…

"When did you get out of Germany, Joe?"

The lugubrious voice: "January. 1936."

So I repeated what I'd said during that long-ago phone call he made to me from Berlin after the publication of *Machines Who Think*. I told him again what I'd said when we re-met at MIT after *The*

Fifth Generation was published and I'd objected so strongly to his trivializing the Holocaust in his essay for *The New York Review of Books*. Now, for the third time, I told him how my husband had got out in January 1939, two months after *Kristallnacht,* three years after Weizenbaum. My husband's every aunt, uncle, and cousin had perished, and his paternal grandmother had survived the French camp Gurs but without her mind intact. I told Weizenbaum my World War II experience. So what was all this about, then? What moral rank did he hold that he might be so judgmental, so condescending to those who believed AI had human promise? This time he heard me.

The face before me crumpled. He *had* depended on that moral superiority he claimed for himself and played to the hilt. In his imagination, he'd concocted a portrait of me as some corn-fed bliss-ninny who had No Idea. He'd cherished his sorrows as unique, superior, inaccessible to someone like me. Now he knew.

I waited for a response.

"Well," he finally said. A pause. "Well." More silence. "Well… Well… Well…"

My parting words were kind, because before me I watched a grave psychological collapse, and it was painful to see.

A day or so later at the meeting, Weizenbaum delivered his usual speech, reminding us all that the German military used computers and therefore. . . A young German artist beside me muttered, "The German military uses knives and forks; let us not use knives and forks." Weizenbaum had become a cruel caricature of himself.

Joe Weizenbaum eventually moved permanently to Germany in pursuit of balm that might heal whatever had broken his heart. Was it exile that had crushed him? But exile from where? From Germany, where he was among millions exiled (and lucky to escape murder) during the Third Reich? From AI, where he'd tried valiantly and failed? From MIT, where an Institute for Computing and Human Values had been established, but he had pointedly not been invited to join? Perhaps the exile was from some imagined paradise that had never existed, and never would. But that's only speculation. He died in Germany in 2008.

By then, whatever beliefs Weizenbaum held of computers being unable to make fine, even humane, judgments were no longer true, even if, back in the 1960s, when he was designing banking systems, it might have been so. In the new millennium, the human race urgently needs help for problems it can't solve with other humans alone, and is slowly, step by step, turning to its intelligent machines for collaboration. The great ethical issues AI raises are getting serious attention, and I'll take that up in a later chapter.

A Long Dance with IBM

1.

In *The Fifth Generation*, Ed Feigenbaum and I were hard on IBM, then the premier U.S. computing company, for its longstanding and very loud aversion to artificial intelligence. The usual story (I heard it first from Arthur Samuel, an IBMer who'd developed the first checkers program in the 1960s) was that T. J. Watson, Sr. worried that loose talk about intelligent computers would scare away customers.

At a Manhattan cocktail party in April 1983, just after *The Fifth Generation's* publication, my husband introduced me to Ralph Gomory, IBM's senior vice-president for science and research, and stepped back to watch the fun. Gomory and I made some polite small talk, and then, also very politely, Gomory said that IBM did not have a corporate anti-AI bias, although it was true they hadn't seen much potential in symbolic reasoning programs until a few years ago. However they'd done "excellent work" on robotics, speech recognition (which he claimed was enough *like* understanding to *be* understanding), and more. IBM viewed the Japanese challenge across a wide spectrum, from devices and packaging to software, including symbolic reasoning programs. In Gomory's words, it was a life and death struggle. He urged me to come up to Yorktown Heights, IBM's major research laboratory, and see for myself, an invitation I soon accepted.[1]

To Gomory, I replied that I was glad for any new information, and it would go into the paperback version of *The Fifth Generation.* But my perceptions about IBM's corporate dogma came from many sources, I said, silent on having heard just weeks earlier in Washington, D. C., from a legislative assistant, who told me IBM had given corporate approval to big government outlays for new generation computers on the condition that they weren't called AI machines. Nor did I bring up (because I forgot) IBM's full-page ads as recently as two years earlier in news magazines and the *Times*, reassuring an uneasy public that computers were only big dumb machines that could never think.

As I stared down into my wine glass, somewhat chastened, a member of the research staff at IBM's Watson Labs in Yorktown Heights named Alan Hoffman barged up to us. He ignored me but said to my husband without preface: "What's all this about expert systems? They haven't done anything so far as I can see. What have they accomplished?" Joe pointed silently to me. "Oh? You're in AI? I don't see any progress in expert systems and their accomplishments are puny." The problems are very difficult, I murmured. Let IBM persuade him, I thought, he's theirs. What his truculence showed was the deep ambivalence among IBM scientists, not surprising. Even the consensual Japanese weren't in total consensus.

A year later, in the spring of 1984, a vice president for systems in

[1] At the same party, I learned (though not from Ralph Gomory himself) that Gomory had given a talk ten years earlier, declaring that AI was the wave of the future. So he meant it. And the program(s) called Watson show his prescience.

IBM's research division—nameless here out of courtesy—gave a major talk at Columbia, and Joe and I gave a party in his honor. He seemed torn, I noted in my journal. Happy about the party—but peeved at me, which showed itself first in a mean-spirited attack on Harold Cohen's computer-generated art on our walls. Our guest was a noted art collector, no philistine.

Having unburdened himself, he asked innocently: "You're not taking this personally?" No, I replied, smiling on the outside, laughing on the inside. "Tell Cohen what I said," he went on. "Oh, I will," I lied politely. Time and again he dug me about the Japanese: "Pamela thinks the Japanese are going to take it all," he explained to the group around us. Pamela believes cooperation, Japanese-style, is better than competition. . . ." "To everything its season," I replied with a smile. Yet he seemed pleased to be honored by this party and thus was a man in minor torment. I was sorry, for I hadn't intended anything personal in the book, yet I saw how wounding it was to him.

I much preferred Ralph Gomory's low-key confrontation: it cleared the air, he offered something concrete as remedy—a visit to the research laboratories in Yorktown Heights—and if I was wrong, there was something I could do about it later, in the paperback version of *The Fifth Generation*, which I did. With this man, I could only shrug.

A month after the art-attack party, Frank Cary, chairman of the board of IBM, was to be awarded an honorary degree at Columbia's commencement. Joe, as head of the computer science department, was asked to escort him; I was to escort Mrs. Cary. With the imminent publication of *The Fifth Generation* in paperback, we thought we'd better send along both Joe's and my CVs, so Frank

Cary wasn't under the impression that Mrs. Cary was to be shepherded by some blameless faculty wife. I imagined Cary aghast, refusing to accept an honorary degree unless someone more obliging (or less insulting to IBM) was rounded up. In fact, he was a model of graciousness, aware but not aggrieved that we'd worked IBM over in our book (an interesting contrast to Gomory, who wanted to take on the issues directly; and the other vice-president, who was peevish without being straightforward). At the morning reception, Joe and I enjoyed talking to Cary so much that we worried that we were monopolizing him.

During commencement, Flora Lewis, the first woman with her own column on the op-ed page of *The New York Times*, gave a splendid talk. But after, I was surprised by the passion it inspired in Cary, who complained, with legitimacy: "You spend years building an organization piece by piece and one of these people comes along and destroys it carelessly…they have their biases, and that's understandable, but they have such power…" I suddenly wondered if he meant me, not Lewis. "Ah," I thought, "though we aren't always right, some of us think very hard about what we write, sensitive to that power." But as Lewis had said, we don't always write what people would have us write about them.

2.

Later in the day, I felt comfortable enough with Cary to speak to him about an issue at the Museum of Modern Art that involved IBM, and my friend, Lillian Schwartz, a celebrated pioneer in computer art and animation. She'd been commissioned to design a poster for the opening of a new wing of MoMA, and it was to be—*ta da!*—computer art, the museum's acknowledgment that, by 1984, computers as a medium for art might actually be legitimate. IBM was underwriting the effort and allowing Schwartz to use their advanced graphics systems, especially their large-scale color printers.

The project had been difficult. For months, the curators rejected everything Schwartz did. First, it "didn't look like computer art"—no jaggies, those stepped borders around images typical of early computer art. She explained that jaggies were being smoothed away by brilliant programming and technology. They complained that the dots were too small to be seen by the naked eye. She explained that the dot-matrix look was also disappearing, thanks also to programming and advanced printing technology.

But as Schwartz digitized and distorted well-known paintings (one of her interim ideas displayed the interior of the museum, a god-like view of the galleries), the curators were horrified that she was deforming sacred art. They eventually picked a New York City scene—she could deform New York all she wanted—but as Schwartz was the first to point out, a straightedge and an airbrush could have achieved the same result. What a missed opportunity this was turning out to be, I thought sadly.

I loved the things Schwartz didn't even submit. For example, she did "homages to" using the palettes of various artists, changing their designs subtly. A grand piece called *Big MoMA* was a six-foot image of Gaston Lachaise's *Standing Woman,* a bold female sculptural figure that often stands in MoMA's garden. Wittily wrapped around *Big MoMA*'s contours were many of the great MoMA holdings: Jasper Johns's *Target* at her kneecaps, Andy Warhol's *Marilyn* at her crotch, Salvador Dali's melting eggs at her breasts, Henry Moore's *Family* in her womb.[2]

[2] Big MoMA had a run of six. One is in the possession of the artist, two in the possession

But all this paled beside the biggest problem. Somebody at IBM stood stubbornly in her way. He was Benoît Mandelbrot, the brilliant mathematician and "father of fractals." Although his reputation was surely secure by now, he was obsessed with the idea that anyone who came near IBM's high-end graphics systems could only be there to steal his fractals without giving him credit.

"Isn't it enough that they adore him at Lucasfilm?" I asked. Schwartz shook her head. Mandelbrot was furious at Loren Carpenter, one of Lucasfilm's great fractals deployers, for what the scientist thought was stealing, not giving credit. I protested this was nonsense: I'd heard Carpenter rapturous in his praise of Mandelbrot.

Moreover, Mandelbrot was soon leaving IBM for Harvard, and Schwartz suspected that to persuade IBM to permit him to take various machines with him, especially an advanced high-resolution printer, he wouldn't allow the graphics systems to be used very much, thereby proving that nobody was using them and thus no one would miss them.

True or not, Schwartz had much material stored on IBM'S advanced graphics machine that she could neither get printed nor have any other access to. Her collaborator at IBM, a protégé of Mandelbrot's, got frantic every time she tried using the machine's editor. She wondered if he'd picked up Mandelbrot's paranoia, that she was trying to steal from Mandelbrot.

Meetings at IBM with the curators (who were having their independent misgivings about this project) always began with

of two of MoMA's curators, and one is mine which I recently gave to Carnegie Mellon University. Where are the other two? The artist, casual about recordkeeping, doesn't know.

Mandelbrot and his assistant already on the machine, tying it up with fractals. Determined to seize the art establishment's blessings on fractals, Mandelbrot, in his hybrid Polish-French-British accent, would launch into an explanation that was well beyond the technical grasp of anyone else in the room. It was irrelevant to the goal of these meetings, which was to view the progress of Schwartz's commissioned work. By the time Mandelbrot was finished, the curators were exhausted and confused, and Schwartz now had to deal with them in their muzziness. This happened again and again, she told me. No matter how early she got to IBM on the day the curators came to visit, Mandelbrot and his assistant already had fractals running on the screens. She didn't know what to do. She reminded me of myself: having this commission was so important to her career that she was ready to be very, very accommodating. As a consequence, she was being run over by both sides. When she'd call me to vent, we always ended up asking each other: would this happen to a man?

On this Columbia University commencement afternoon, I gave Cary the briefest possible précis of the situation. At his Parnassian level, he hadn't known that IBM was underwriting the poster commission, but, as luck would have it, he was not only IBM's chairman of the board, but also on MoMA's board of directors. He nodded, took names and numbers, and sliced through the problem in two days. Lillian called me gratefully to tell me.

3.

In short, Ralph Gomory was right. Despite years of advertising to the contrary, IBM took AI very seriously. The firm's successes in the late 1990s and in the 2000s were delightfully public and decisive—Big Blue's defeat of the world's human chess champion, Garry Kasparov, and Watson's triumph in *Jeopardy!* In early 2014, IBM announced that it was investing a billion dollars in machine intelligence, and that October the Watson Research Group moved to the East Village in New York City. Watson was already at work on real-world problems in medicine, scientific research, management and sales guidance for large and small businesses, even on teaching devices disguised as toys.

For example, in partnerships with medical centers, including the Mayo Clinic, the Cleveland Clinic, Sloan-Kettering, Baylor College of Medicine, and Columbia University Medical Center, Watson scrutinized patient data to guide better outcomes for cancer, such as genomic implications, or faster matches of patients to appropriate clinical trials. Watson has been identifying the proteins associated with cancer, a search that yielded one per year in the old days, but Watson was finding them at the rate of seven per year. These then suggest new targets for chemotherapy. "Watson has truly become a colleague to clinicians in making treatment decisions," said Lauri Saft, director of IBM Watson Ecosystem (Morais, 2015).

With what IBM was calling cognitive computing, Watson was learning to think like humans think. Rob High, chief technology officer of IBM Watson Solutions wrote:

> These machines won't be our adversaries. Instead, they'll augment our knowledge and creativity with skills that they're really good at, including computation, memory, speed-reading, and the ability to find insightful patterns in huge quantities of data. Computers will be our ever-present intelligent assistants. Thanks to cloud computing, a wide variety of software programs called cognitive advisors will be at our beck and call whenever and wherever we need them….Cognitive machines will democratize expertise. (2013)

Distribute it, anyway.

At the Tribeca Film Festival in New York City in 2015, Watson illustrated these kinds of help. IBM's Lauri Saft told the audience:

> Film and artists and creative people and narratives—that is the essence of what Watson handles best. Words and language and sentiment and ideas, right? That's what Watson does for a living.. . . It's man with machine, not man versus machine. There are things that we do really well as humans, but there are also things that systems do really well. (Morais, 2015)

Watson could be a colleague to help screenwriters with ideas, with plots, with combinations of traits for characters. Saft told Betsy Morais of *The New Yorker* that "He's constantly saying, 'What about this? We could do that,' perpetually feeding you with ideas" Watson will not unseat Steven Spielberg: "You need that combination: people and machines are more powerful together" But to Saft, Watson was already *he*.

In Chapter 14, I described IBM's Project Debator, meant to be another form of personal assistant. Meanwhile, along with being a lab research partner, clinical colleague, financial advisor, and a collaborator in the arts, Watson published "his" first cookbook, *Cognitive Cooking with Chef Watson*. If you can tolerate twenty-five-step recipes, these concoctions will wow your dinner guests.

Analysts complain that Watson is losing money and isn't that good anyway. Scientific tides ebb and flow. But I see Watson as an amazing about-face from the days when old T. J. Watson worried that even the mention of intelligent computers would scare away the customers. Now, expert systems are for every one of us. We've come a long way since Grandpa Dendral.

Being a Nine-Day Wonder

1.

Back in the early 1980s, the field research for *The Fifth Generation* was eye-opening, but more so was the aftermath. The book became a best-seller in Japan and the United States and eventually, with authorized and unauthorized copies, sold about half a million copies around the world. Ed Feigenbaum and I enjoyed the giddy experience of being nine-day wonders. Yes, it's fun to walk along Madison Avenue and see your own book in bookshop windows. Translations abounded, the phone rang constantly, and the publishers and we were happy at last. Congress held hearings: was Japan's Fifth Generation a threat to national security? Should the National Science Foundation or DARPA invest more dollars in computer research? Snarky researchers claimed that Feigenbaum and I had written the book only to beef up his research budget. In that farfetched scenario, I was a willing and invisible tool. But I was to become invisible in other ways, too.

Invitations poured in to be interviewed, give talks, be lionized, be attacked. Among the odder lionizations for me was a phone call from *Vogue* magazine. They'd seen me interviewed on public television's *MacNeil-Lehrer NewsHour* and wanted to conduct their own interview. Because they'd fingered me on a relatively serious TV news show, I thought their interview would be moderately serious, too.

But no. What did I think about clothes? (I loved them; I cared about self-presentation, as anybody who's seen me knows.) What was my idea of a balanced day? (Being left alone to work.) Did I see enough of my husband? (Sure. Next question?) Work comes *first*? My goodness! And your husband knew that before you got married? Goodness! Finally, if I could throw it all over and escape to paradise, what would paradise look like? I thought I was already in paradise in New York City and harbored no urges to open a bed-and-breakfast in New Hampshire or a boutique in Mendocino.

A few months later, I picked up a copy of *Vogue* at the airport and saw I'd been placed in an article called "Life at the Top: Women Talk about Success, Time, Love." I was amused, especially as I sounded like the compleat nerd. But I was sandwiched between Alice Waters and Mrs. Byron Janis, an apotheosis of some kind, I guess.

2.

It pains to me to say that a best seller also brought out the worst in journalists. Not my first experience: The *Newsweek* cover story for June 30, 1980, was "Machines That Think." The issue appeared just weeks after *Machines Who Think* was published but had no mention of my book or me. Possibly coincidence.

When *The Fifth Generation* was published, *Newsweek* made it a cover story, pictured my coauthor, mentioned the book—but never once mentioned me as coauthor. "Maybe the third time I hit," I wrote in my journal, "I'll even see my name attached to my ideas." More egregious was *Time* magazine. Writing about Japan's effort, the journalist carved out some of the best parts of the book and wrote as if he'd discovered them for himself. Whole quotes were lifted from the book and attributed solely to Feigenbaum, but neither the book nor I was mentioned. To this I did write a note of protest—polite, slightly humorous, saying if writers don't respect each other as specialists, who will? The theft is more than petty.

The journalist called me immediately, deeply apologetic. He'd done it, knew he shouldn't have, and was even more ashamed to receive such a nice note from me instead of the chewing out he richly deserved."You were so quiet when you were in my office I figured you didn't have much to do with the book." To that I was nearly speechless. Did he think I was Ed's NYC bimbo? No, what he saw was years of socialization, of being properly deferential to men. I told my sister the story, and she said: "I expect you've learned a lesson you'll never forget." I reported this in my journal and then went on to speculate:

June 28, 1983:

> *But this raises interesting questions about how writers are perceived, especially by other writers. I'm put in mind of the reporter who picked my brains for fifteen minutes then asked me for the name of an "expert" who could tell him all that. In some sense, it reflects American anti-intellectualism, which is to say that we honor doers, no matter how good or bad, but not those who think about things and try and make sense of them. But why should journalists be the worst at that? Interestingly, people in the field I write about don't have that attitude. To them, I'm a professional—in a different profession. They appreciate how well I perform my craft, and scold me when I falter. I'm not interested in celebrity, but I am saddened by lack of public recognition, which is slightly different. And that I lay to the curious attitudes of members of the press.*

> *Speaking of which, one such called me a day or so ago to ask questions. He'd obviously read nothing by me (nor anyone else in the field). How can people waste your time like that?*

May 21, 1984:

> *Joe hears from David Lee, his former graduate student, that David met an IBM researcher at Yorktown Heights who's just come from Stanford. "How come," says this young former Stanford student, "that Pamela McCorduck let Ed Feigenbaum put his name on that book when she could just as easily have done it herself? Everybody at Stanford knows Machines Who Think. Everybody knows that Feigenbaum not only didn't write The Fifth Generation, but hardly knows what's in it. Why'd she do it?"*

The IBM researcher continued with further wicked graduate student folklore, each tale growing taller and more distant from anything resembling fact.

But I'd endured so much "only the ghostwriter, not coauthor," even from the book's original publisher, so much "we only want to interview the expert," that I laughed. Hard.

3.

Feigenbaum and I were invited back to Japan from time to time to celebrate Japan's coming of age in computing. Tokyo, November 8, 1984:

> *That evening, the banquet. Go to pick up Ed, who sits on his chair, barefoot, and says, "I can't go on. I'm so tired. I just can't go on." I say,"Cheer up, you'll make it. Only two hours and then you can go to bed." We go down and get into the reception line, where Ezra Vogel joins us.* [Vogel a well-known Harvard expert on both Japan and China.] *We three are chatting briskly when two kimono-clad ladies walk up to the men, bow, say "spesha guest" and lead each*

> *of them away, leaving me standing there looking more than a bit forlorn, I'm sure. Furokawa-san has been watching this, rushes over to me with a "spesha guest" chrysanthemum tag, for sure enough, I too am a "spesha guest," but the ladies are so unused to a woman being such that they've overlooked me. An elegant French-style banquet, and I'm asked to say a few words. I talk about the momentous occasion not only for Japan, but for the human race. Had I known I was one of only two people to make such remarks, I'd have worked harder at it, but short and sweet, thinking I'm one of many. Afterwards, much to-ing and fro-ing, and Ezra, Ed and I retire to my room, where Ed, risen to life like the phoenix, holds forth until Ezra literally falls sideways off his chair in jetlagged fatigue, and I'm incoherent. This is the man who could hardly put his shoes and socks on when I went to get him.*

November 11, 1984*:*

> *Ed takes me to Akihabara, where I see more electronics than I'll ever want to see again, and demonstrates something called a karaoke machine, standing in the shop, singing "I Did it My Way," for me, except he doesn't know too many of the words. I didn't know who I wished could see us then, but I certainly wished somebody we knew had ambled by.*

In April 1983, just after *The Fifth Generation* was published, I was invited to a Washington meeting to help plan a film for the U.S. Pavilion at the upcoming World's Fair in Tsukuba, Japan, to be produced by the United States Information Agency (USIA). Japan's dramatic, and to Americans, shocking leap ahead in AI, must have a response from American AI, and a number of AI researchers were gathered, most of whom I knew.

In the air, I sniffed a strange piety. The USIA's brochure for American businessmen had stressed competition, but my AI acquaintances were speaking somewhat sanctimoniously about cooperation instead. I laughed to myself, wondering how much "cooperation" existed in their home institutions. Did they share their DARPA grants with the starving English department?

The meeting continued over dinner. After a while, Roger Schank, then at Yale, began a vitriolic attack on Feigenbaum, not a surprise. But Harry Pople at the University of Pittsburgh began to attack Feigenbaum too, the gist being that he was overselling expert systems and raising expectations that could never be met.

Finally, I decided I'd better say a few words on behalf of my friend and co-author. This came as a surprise to Schank, who seemed embarrassed; a relief to David Hertz, sitting beside me, who said he wondered when I'd speak up. I'd been silent because I was amused, because I was loath to get into confrontations over dinner, because the spectacle was so interesting, and because I wanted to see how far they'd go.

I began by saying that Feigenbaum had a right to whatever claims he was making about expert systems, and furthermore, the expert system he praised most generously was Harry Pople's, giving full credit to Pople everywhere he went. Pople looked a little abashed (but not as much as he should, I thought) and sputtered some inane defense. Then I turned to Schank. "You're marketing these things. Are you marketing a fraudulent product?" "Of course not!" he cried.

I went on to say that in our book, Feigenbaum had been the cautious one, toning things down. "Ah," said Pople, "perhaps there's a difference between the public persona and the private persona." It was useless to point out that nothing is more public than a published book, and I changed the subject. I happened to talk to Raj Reddy right afterwards and confided all this to him. "Professional jealousy, nothing more," he said, confirming my own feelings.

This kind of sniping had many faces. Tony Ralston, a computer scientist at SUNY Buffalo and the editor of a short-lived journal called *Abacus*, wrote to ask if I'd do a profile of Feigenbaum. I declined because my friendship with Ed was so close and enduring that I couldn't even pretend to objectivity. My views, I added, would have to wait until my old-age memoirs. "But they'll get a hack," I wrote in my journal, "who'll miss the marvelous range and breadth, and listen to his jealous dwarf enemies."

To tempt me, Ralston had sent me a review written by some nonentity who presumed to review the oeuvre of Donald Knuth, a giant of computer science, whose *The Art of Computer Programming* is one of the momentous scholarly achievements of the field, the multiple volumes known deferentially as Knuth: "Look it up in Knuth." Although the review was mainly praise, the discrepancy between Knuth's accomplishments and those of the reviewer was so wide as to be ludicrous. It made even praise of Knuth sound presumptuous. To add to my irritation, Ralston added I "ought to know" that this month's *Abacus* had a somewhat harsh review of *The Fifth Generation*. Why ought I to know? Should I un-write the book?

4.

Expert systems were in fact problematic. Ed Feigenbaum and his colleagues had upended AI research, pretty much devoted earlier to games, mathematics, puzzle-solving, and some attempts to model modest instances of human problem solving. They'd insisted on putting real-world knowledge into such a system, so that it made decisions the same way a real-world, intelligent human expert might. By taking such a step, Feigenbaum himself left the kind of AI he'd been trained in, cognitive psychology and the modeling of human memory, and moved to what would later be categorized as a knowledge-intensive rational agent. He was less interested in emulating human cognitive behavior than in producing a system that would perform as well as, even better than, human experts in the real world. He'd had some stunning successes and was plenty excited about this brainchild of his.

Expert systems were difficult to build and almost as difficult to maintain. Each application seemed to draw smart young graduate students into pushing the field of AI into one more app, instead of pushing it forward, beyond this relatively fragile early paradigm. It was a path of easy payoff and least resistance.

But as I heard Feigenbaum's colleagues blame him for a waste of brainpower and other resources, I was puzzled. If a better path lay elsewhere, why weren't they taking it? Or more important, guiding their graduate students to take it? Why weren't they persuading the funding agencies that AI had a different, better future? Why blame businesses for taking up this technology, which seemed on its face useful to business? Why blame Feigenbaum and fail to take responsibility for moving the field forward themselves?

5.

Eventually, the Japanese Fifth Generation did not reach the ambitious goals it originally set. Chapter 18 offers some possible reasons why. But the Fifth Generation's major accompishment was not negligible: it trained a generation of young scientists in the field. Thus Japanese AI research continues apace. Yet an unfortunate chasm has appeared between Western and Japanese research. Ed Feigenbaum thinks the biggest wedge is the language barrier. Westerners fail to learn Japanese, and Japanese scientists are less and less interested in learning

English. Perhaps neither side thinks it can learn much from the other. This is a great pity, and I hope the chasm closes.

Time passed. I wrote another book, *The Universal Machine,* and was mulling a new book on how people adapted to great changes in their lives. Feigenbaum stepped into this mulling process and suggested we collaborate again. This time he wanted to include Penny Nii because the book would be about expert systems in business, and she knew as much about them as anyone. By then, Joe and I had moved to Princeton and bought a drafty old house in desperate need of new windows on two floors. We faced an outlay of tens of thousands of dollars, and we were already financially stretched with the house itself. This new book offered me chance to see if I could write a good book about business and certainly a chance to earn some much-needed money to make the Princeton house habitable in the winter. Above all, because *The Fifth Generation* had been so much fun, I couldn't see any problems.

Expert systems had burgeoned in universities and in firms. Hundreds of them now embodied the knowledge of experts in science, business, medicine, and many other fields. Each system's knowledge was explicit and offered ways of communicating, exchanging, and improving on it. The disadvantages of that sudden growth were a premature commercialization of AI (fragile systems that were difficult to maintain) and, as some argued, a lateral instead of a forward push in research, as each little piece of expertise was shoehorned into an expert system and earned somebody, somewhere, a masters' degree.

Feigenbaum, Nii, and I began work on the book with a marathon round of interviews in firms that were using expert systems. One interview took place in the Houston offices of Schlumberger, the oil field services firm.

March 12, 1987:

> *At a Schlumberger installation in Houston to interview. Praise the lord, Schlumberger is run by the French, so lunch is the first decent meal in Texas. Simple but excellent. Lunch is also info-gathering for me: I listen to the Houstonians (even as I once listened to the Japanese) gather pearls from Ed The Guru. The truth of the matter is, The Guru is pretty damn smart, so they're correct to regard him with reverence. That combination of native intelligence and sheer decades of experience is priceless. I think of Ed as on top of things in every sense. Now, why can't I get excited about expert systems? Partly they lack the plethora of grand ideas I'm used to. Partly it's a writer's problem, trying to make the stories appear different when they're basically all alike. But that's an old storyteller's problem.*

Buried in that book we eventually wrote, *The Rise of the Expert Company*, was a subtle prophecy. Ed Mahler, a senior chemist at DuPont, was high on expert systems and had formed a group to introduce them in DuPont's various branches. But Mahler thought you didn't need fancy programs, fancy machines, or knowledge engineers to turn it all into code. You gave a chemist (or another scientist) some means of interrogating himself, you gave him a laptop, and you made him his own knowledge engineer. That's pretty much what happened, upending many a business plan that proposed to manufacture special-purpose machines for AI or to train knowledge engineers to go across disciplines, like business consultants, evoking knowledge from human heads to be cast into computer programs. Later, machines would begin to develop their own expertise, learning from the environment, whether they did it supervised by an expert or independently. The idea animating expert systems became fundamental, an implicit part of a much larger genre of programs called knowledge-intensive rational systems, whether that knowledge came from huge data sets (comprising aggregate human behavior such as your Facebook loyalties, your response to digital ads, and your smartphone conversations) or directly from human heads.

Ed and I kept talking to each other as always, but I began to take a rest from AI. The field seemed to have passed from revolutionary to normal science, in the sense Thomas Kuhn means in *The Structure of Scientific Revolutions*. For me it was no longer as interesting. Change would come again, but until then, I found other things to explore.

Breaking and Entering into the House of the Humanities

1.

I'd given up the idea of writing a biography of Herb Simon and returned the advance to the publisher. I was just too close to write anything but hagiography, and he deserved better. I wasn't idle: commissions appeared from the dozens of magazines that suddenly blossomed in the early 1980s to present science to an apparently insatiable lay readership. I was teaching science writing at Columbia University and also worked for women's magazines, *Cosmopolitan* and *Redbook*. They knew what they wanted, they gave expert editorial guidance, we all had fun, and I made money. But journalism, with the exception of *Wired*, was basically frustrating,

I wrote another book, *The Universal Machine,* a series of connected essays about the worldwide impact of the computer. Although both my agent and my editors thought the book was good, it fell into the hands of a self-declared humanist, who reviewed it for *The New York Times,* hated computers (and by extension me), and shellacked it. The review was one of the few times that this kind of ignorance didn't make me laugh. Deep points in that book were worth making.

However, the reviewer was a part of the First Culture that remained deliberately ignorant of what was already possible and impervious to what lay ahead.

September 13, 1984:

Lunch at the invitation of the chairman of the English department at Columbia. In his opinion, computers might possibly be useful, but his colleagues are hostile. The usual reasons for this, including "another passing fad" which stuns me into silence. What he really wants to talk about is that he feels his PhDs are at a comparative disadvantage for not knowing word processing. Can I suggest anything? Lobotomy, I think, smiling politely all the while.[1]

About this time I was at a dinner party with the Nobel Laureate physicist I. I. Rabi, who leaned over to me with a good-natured chuckle and said, "You can learn a lot from the humanities." Pause. "But not from the humanists." I often walked Rabi up the hill from Riverside Drive to the Columbia campus (he had a well-calculated route to avoid the ferocious wintertime winds that blew along 116th Street) and I once asked him how physicists reacted when he brought back from Germany in the 1930s all these new-fangled ideas on quantum physics. How long did it take them to accept the new? He laughed merrily. "Never! I had to wait for them to die."

[1] In 2012, Harvard University, never an institution to rush precipitously into change, released a report on revitalizing the humanities at Harvard. The report's pervasive theme was the imaginative use of the computer. Granted, this occurred some thirty years after my lunch with this particular English department chairman. At Harvard, plunging enrollments in the humanities had helped inspire this reevaluation.

2.

In *The Fifth Generation*, I told a story recounted to me by numerical analyst Beresford Parlett and worth repeating:

> It was early in July 1953, a rare hot day at the end of the summer term at Oxford. Two punts were being languidly poled down the river Cherwell, filled with high-spirited young men who were on their way to a twenty-first birthday picnic for Beresford Parlett. Parlett, who would later become a professor of computer science at the University of California, Berkeley, was an Englishman with an affinity for American friends, and it happened that his punt carried the college's American contingent of Rhodes Scholars, men who were studying economics and mathematics. Among them was Alain Enthoven, later Assistant Secretary of Defense for Systems Analysis and still later, a professor of economics at Stanford University. Enthoven stared meditatively at the punt ahead of them, which contained, by everyone's estimate, the brainiest young men in the college. They were all "reading greats"—studying the Greek and Latin classics. "There," said Enthoven, fixed on the punt ahead of them, "*there* is England's tragedy." (Feigenbaum & McCorduck, 1983)

I reported this then because it fit neatly into the saga of British efforts in AI. But I missed its fundamental significance. Accidentally or by design, in the early 19th century, universities had become a ghostly simulacrum of the British class system—*belles lettres* at the top, the study of music and painting, history and philosophy just below, and so on, all the way down to the contemptibly practical, like science and engineering, considered no better than intellectual shopkeeping.

Here's another way British education preserved the class system: My father, a clever boy, had left school at fourteen, as most working class children did in 1926. A few years later, when the Great Depression had already arrived in Europe, he sat for a university scholarship and, to his deep joy, came in first in the competition. The authorities congratulated him warmly but then told him that of course the scholarship must go to Lord So-and-So's son, who could actually use it.

For me to suggest a half century later that artificial intelligence—whose very name put people on edge—might have something to do with the mind, might profitably be attended to by people whose brief was the human mind, was hopeless. AI was about machines and engineering. One might as well suggest a wedding to the dustman. Although the gods of the First Culture continued to reign in Valhalla, their unwillingness to consider the digital world had already put the castle to the torch.

In mid-April 1985, my agent sent me to talk to various editors about ideas for books. One of them, occupying one of the loftiest thrones in Valhalla, spent a while with me, assuring me he was an agnostic on the subject of the information revolution (though he lamented the millions his firm had spent on teaching programs instead of textbooks). He asked me—justifiably—are things changing? Isn't it really that expert systems will only replace people who weren't really experts? I replied with the cost effectiveness argument: at that time, no firm would build an expert system except to replace a costly expert—the undertaking was just too expensive. "But really," I wrote in my journal, "it's just the old 'if it's intelligence, it can't be automated' argument."

The culture clash was acute. As I'd entered, he showed me proudly that Joe Weizenbaum had contributed a blurb for one of his authors, and I paused to think anyone took this seriously. Yet because I was so awed by this editor's name and splendid eminence in the publishing world, I was tongue-tied trying to explain myself. Afterwards I understood it was also because the paradigm had shifted for me. It hadn't for him. It was useless to say that.

Although I was intimidated by this editor's First Culture renown, I could also see how pompous he was (pontificating at length and so softly that I could barely hear him over the air conditioning, deliberately causing me to lean forward over his desk), how oblivious he was to the intellectual excitement surrounding the computer. He fit perfectly C. P. Snow's old description of the First Culture: he was an "intellectual" and excluded from that category anything that didn't interest him. Culture meant only what he said it meant. An atavistic reverence from my youth had overcome me, that younger part of me that honored and wanted to be accepted in the First Culture. Part of me mocked myself; part of me wondered why I longed to be welcomed.

> *For this major editor was like the minister of culture of a tiny, once important country, who hasn't yet had the news that power has shifted. God only knows what he thought of me, gasping and burbling, totally undone by the border crossing. He thought: here's another techie who can't put together a declarative English sentence. I thought: once more I question the worth of language, heresy for a writer. Weizenbaum as an endorsement to be proud of? World literature teems with the pious hypocrite, from Tartuffe to the Reverend Arthur Dimmesdale to Uriah Heep: what's the point of being the Grand Vizier of Literature if you can't detect one on the hoof? "The mustard gas of sinister intelligent editors," Allen Ginsburg had written in Howl.*

Said my husband consolingly afterwards: yes, you'll have to wait for this generation to die. The next generation will wonder what all the fuss was about.

I, Sisyphus. But as Camus argued—long story—Sisyphus was happy. Me too.

In the next thirty years, the heated (or pleaful) arguments for studying the humanities arrived as the night follows the day. The first point was nearly always that the humanities teach critical thinking. Did anyone doubt that studying to be a scientist or an engineer requires sharp critical thinking? Clarity of expression, then? No: that's the purpose of composition courses, left routinely to teaching assistants, adjuncts, and specialized remediators. Reading novels and poems trains your empathy? More like it. The humanities enrich your life? You bet. Profoundly. But the world had changed: students needed to know they'd graduate with some promise of gainful employment to pay off the staggering debts they were now incurring. For that, the humanities seemed unpromising.

3.

December 21, 1985:

> *My survey of contemporary American literature tells me one could read a sizable chunk of it and be innocent that any technology besides the telephone and the internal combustion engine affect modern life. Since both these are a hundred years or more old, this doesn't seem especially brave on literature's part. Norman Mailer, writing to invite PEN members to the coming international meeting, jokes that no one pays attention to writers. But why should they? Meanwhile, the writer's imagination, the imagination of a religious fanatic, believing divinity on its side, is fat and megalomaniac, dreaming it has answers. Preposterous: it doesn't even know the questions.*

The following month, Joe and I were in California. I'd been invited to join a panel at Santa Clara University in Silicon Valley to respond to the remarks of the main speaker, Ashley Montagu. A celebrity pop anthropologist, forgotten now, he'd made his name with twenty popular books and was a fixture on late-night TV talk shows. Hundreds were turned away from the large lecture hall.

January 11, 1986:

> *My worst fears are realized when I hear, first, that he always talks by "spontaneous combustion," as he puts it, and second, that he long ago stopped reading other people's books. He artfully quotes Hobbes, that if he were to read the work of others, he would be as ignorant as they. I judge he's a man living on his intellectual capital, and I'm right.*
>
> *He tells us the topic—the impact of the computer—is so important, however, that he's going to read his speech, which he doesn't. When we're fifteen minutes along, and still on The Fall (with Cain and Abel thrown in) it looks to be a long night. I'm fascinated by his technique—many irrelevant parentheses, mainly jokes at the expense of the professoriate and other professions, snatches of poetry, storytelling—and fascinated too by how he evokes sheer adulation from the audience, as if their critical faculties were simply nonexistent. I have a pretty dilemma. To tell the truth, thereby making enemies of the seven hundred who adulate, but also permitting me to wake up and face myself tomorrow morning; or to be a well-mannered guest. Eventually I choose to praise him for his truths—though platitudes most of them are—and merely "raise questions to which I have no answers" about some other topics he's raised.*
>
> *For instance, if we have evidence that people have been dehumanized since the agricultural revolution by their technology, maybe, after ten thousand or more years, we need to redefine what it means to be human. I add that since (as Dr. A.M. has correctly pointed out) tools are human thought made manifest, then it can't be dehumanizing to come face-to-face with another aspect of our humanity in the computer.*
>
> *And so on. I put in a plug for computer science, which A.M. obviously doesn't understand, but feels free to criticize. I say more, including raising doubts that*

> *happiness is the same as the simple, untechnological life. (A.M., meet Raj Reddy.) But all the time I was immensely polite with my Nice Girl smile, and kept congratulating him for his insights. I said not a word at how shocked, and then contemptuous I felt of the audience, that this string of platitudes, half-truths, and outright fabrications were so inspiring to them. Truly the triumph of style over substance, and frankly, I could take a lesson.*

November 8, 1987:

> *At the Art Institute of Chicago. A perfectly awful panel led by some young woman at the Art Institute School, who misunderstands computers, art, and, God knows, physics. She repeats from time to time: "Here's how I FEEL about physics…" The artist Harold Cohen beside me is snorting in rage. I'm laughing. Yes, dear, please tell us how you FEEL about physics. I was put in mind of 19th-century Margaret Fuller's apocryphal declaration: "I accept the universe!" and Thomas Carlyle's reply: "Egad, she'd better."*

4.

Yet all the while I kept asking myself: was I so high on what promised to be one of the grandest intellectual accomplishments of humanity that I was cruelly impervious to the deep, perhaps unconscious, fears of the humanists? Did I fail to see how frantic they were that the earth was shifting under their feet—or the smoke was drifting upward from Valhalla's cellar? Couldn't I moderate my enthusiasm, extend empathy, compassion?

No. They were neither fearful nor frantic. They didn't need my compassion and would have refused if I'd offered it. They were the aristocracy, sublimely self-assured in their faith.

That then raised another question: did they not actually assimilate the texts they claimed to honor? The texts that, over the centuries, counseled open-mindedness and humility in the face of the new; counseled caution in the arrogance of faith; texts that mocked the complacency of the status quo, that ridiculed the zealously pious (who always had a seamy underside—celestial thoughts and subterranean conduct, as Montaigne put it)? Did they draw no lessons? Learn to tell the ersatz from the genuine? *Did they not for a moment think that this could be important?*

Yet some humanists were intrigued. I described a holiday party we'd attended our first semester at Columbia:

December 24, 1979.

The humanities professors—French, history, philosophy—showed how the first-rate are so different from the second-rate. A) They're fascinating to talk to on their own subjects, having a wide and relaxed view, and B) they're eager to know about other things, and welcome news about whether artificial intelligence will have an impact on their own field; very little in the way of derisive laughter, or the just plain indifference I was used to getting from the English department in Pittsburgh. Or maybe it's what my mother would've called breeding—no matter how ridiculous you consider your conversation to be, you politely dissemble. Either way, more than a few cuts above what I'm used to from the humanities.

November 11, 1986.

Two days at Kenyon. A feeling of letdown. I think: at last, recognition, but it's the techies who've invited me. The English department—largest by far on the campus—doesn't know what to make of me (nothing, in the end: the chairman shifts uneasily as we're introduced. "Oh yes, I'd heard, uh, maybe we could get together tomorrow?" this vaguely, and doesn't come to my *lecture so I think to hell with him too). The students are marvelous—sharp without being smart alecks (though the women still keep silent; I have to draw them out). Very much like the Chautauqua experience* [I'd given a talk there the previous summer]. *I both admire and am appalled by such hermetic decency. I'm too hard—the artists came to my talk and were enchanted. The techies were grateful I'd come to validate their worth. When I give Joe a précis tonight, he understands, says once more: yes, and when the dust settles you'll get no credit. . . .Reading Furbank's* Forster. *E. M. Forster understood he was "important" by his mid-20s. I spend my days reconciling myself to my unimportance, hoping against hope it isn't so.*

Septmber 6, 1988

I accept an invitation to speak on AI and the Humanities at Pitt. The enthusiastic organizer tells me he's going to get right off the phone and "tell everyone you're coming." I laugh, do not tell him how the book he's praised most, MWT, was the cause of my banishment from Pitt. Ah, the wheel always turns. . .

As I recall, no one from the Pitt English department came to that talk either.

5.

History loves irony, and so the 1980s were exactly the decade that AI research was moving brusquely and impatiently into territory long claimed by the humanities, particularly philosophy. What was mind? Could it be, as Marvin Minsky proposed, a "society" of competing, relatively independent agents inside your head, each one jockeying for dominance? If you moved from representing *problems* to representing *knowledge* in a computer system, just how was this knowledge to be represented? Was a general-purpose representation possible, or did different kinds of knowledge require different representations? If you chose a general-purpose representation, how did you organize and connect knowledge in several domains? After you chose a suitable representation, how was the ontology, the agreed-upon knowledge, to be kept consistent and valid? How were beliefs, or even truth, to be revised, validated, and maintained in the light of new knowledge?

Philosophers from at least Aristotle on—including more recently, Charles S. Peirce and Ludwig Wittgenstein—had wrestled with these issues with little success. AI was quietly breaking and entering into a lordly old mansion owned by the philosophers for centuries. Unfortunately for AI, most of the rooms in that mansion were vacant.

For decades, those few philosophers who considered AI even worth their attention treated it like a great game of poker—grave visage, dazzling plays, strategies, bluffs, and quick adaptation as the game changed. Every once in a while, a fellow philosopher named Daniel Dennett would stop by this game for a few hands, clean out the pot, and depart.[2] The other players hardly noticed. After all, what mattered were bravura playing and clever rhetoric for each other and especially for the partisan spectators ("I *knew* machines could never think and now you've proved it").[3] No. They'd invented parables, but hadn't proved anything.

[2] Bruce Buchanan, a principal of the Dendral program and other pioneering AI work, had certainly earned his PhD in philosophy, but he'd gone over to the dark side so early that people outside the field hardly considered him a philosopher. "I wanted to do something important," he once told me. And so he did.

[3] It bears repeating that Daniel Dennett and I nearly always end up in the same place, but he does the heavy lifting of thinking us through to that end, while I arrive by shortcut. See especially his 2017 book From Bacteria to Bach and Back: The Evolution of Minds (W. W. Norton).

In its intellectual contributions, the philosophers' game was finally inconsequential and embodied the Arab proverb: the dogs bark and bark, and still the caravan moves on.

AI had a problem that philosophers had never faced: its researchers needed to write programs that demonstrably *worked*. Forced to make vague concepts precise enough to turn them into executable computer programs, researchers of the 1980s were absorbed with figuring out more of what constituted thinking: programs that planned ahead and took into account limited resources, such as time and memory. Programs began to learn from explanations and began to function in an environment where multiple agents, often in conflict with each other, needed to act.[4] During this decade, foundational work in applied ontology emerged because truth maintenance was suddenly a necessary goal, a way of keeping beliefs and their dependencies consistent. Systems cleverly increased the speed of inference and exhibited a much better understanding of the interaction between complexity and expressiveness in reasoning systems. Artificial agents began to use psychological reasoning about themselves and other agents.

All these sound dauntingly technical. They are. Not then nor later did they lend themselves to sexy journalism or inspire Dionysian passions. Indeed, those years were sometimes described as "the AI winter," largely because no one could figure out how to monetize such research. But the work is the anatomizing of what, for centuries, was casually known as intelligence—along with all its synonyms: cogitating, reasoning, considering, planning, keeping consistency,

[4] One early cooperative multiagent program was "boids," a program that simulates the emergent behavior of flocking. Its creator, Craig Burton, eventually won a special award from the Motion Picture Academy of Arts and Sciences for the program's application in such movies as Batman Returns.

inferring, leaping to conclusions, drawing parallels, imagining, mulling, analyzing. *Intelligence* is a suitcase word, in Marvin Minsky's phrase, a word that needs careful unpacking to reveal all it contains. Moreover, revelation isn't sufficient. Each part of this deeply complicated process must be understood and then described in explicit detail so that a computer can carry it out.

I've said AI was doing normal science, in the Kuhnian sense of normal, as distinct from revolutionary science. It was dynamic and abundant nevertheless. Those advances would raise further challenges: as data sets grew larger and computation faster and deeper (but more costly in both time and computational resources), how could searches that could never be exhaustive instead be automatically guided? How could goals be reached in a timely way? This was exactly the quarrel Herbert Simon had earlier with classical economists and their impossibly idealized Rational Man, who could never explore all alternatives to arrive at a rational economic decision. Searches needed to be guided, and tradeoffs made between computation costs and timeliness. Meta-level reasoning, over and above the busy lower-level searches, had to find those balances and make those tradeoffs in real time. These were tremendous, exhilarating challenges for AI researchers then, and they remain so now.

Earlier, Marvin Minsky had quietly said to me, look how long it's taken physicists to get where they are. Surely intelligence is as difficult as physics. Martin Perl, the Nobel laureate in physics reminds us: "The time scale for physics progress is a century, not a decade. There are no decade-scale solutions to worries about the rate of progress of fundamental physics knowledge" (Overbye, 2014). Intelligence is at least as hard, at least as exhilarating.

Indifferent at best, usually hostile, the First Culture disdained it all.

Part Five: Silicon Valley Sketchbook

I have perceiv'd that to be with those I like is enough,
To stop in company with the rest at evening is enough,
To be surrounded by beautiful, curious, breathing, laughing flesh is enough,
To pass among them or touch any one, or rest my arm ever so lightly round his or her neck for a moment, what is this then?
I do not ask any more delight, I swim in it as in a sea.

—Walt Whitman, “I Sing the Body Electric”

The Silicon Valley Sketchbook

1.

Jaron Lanier is a striking looking man by anyone's estimate. His ginger-colored dreadlocks always reach at least to his shoulders. He so resembles the self-portrait of Albrecht Dürer, painted at age thirty, that from Munich once, I sent a postcard of that painting to Lanier just for fun. When he was living in New York City, he sometimes had trouble getting cabs to stop for him because of his appearance. If he, Joe, and I went out to dinner together, afterwards I'd save us all trouble by stepping out to hail a cab—any cabbie will pick up a middle-aged white woman—and then give Lanier a big hug, and open the taxi door for him, to the driver's astonishment.

Under those dreadlocks, Lanier has a sweetly cherubic face, the celestial effect enhanced by his habitual black or white attire. That cherubic face reflects a kind, gentle, and deeply decent soul within.

Jaron Lanier also has one of the most interesting minds in computing. The conventional wisdom is never Lanier'swisdom. You might not agree, but you're always stretched to argue.

We met first in the summer of 1985 when Lanier was chief scientist of a startup he'd founded called VPL. He confided the initials "sort of but not really" stood for virtual reality, a term he's credited with popularizing. The firm was selling systems that produced make-believe reality, realized electronically.

For this, I donned a headpiece with tight-fitting goggles. Before my eyes, an electronic landscape appeared that Lanier kept assuring me he'd knocked together over the weekend, a Greek-like temple in a pastoral landscape. I wore a glove that let me interact with this landscape. Heights scare me, so I particularly remember being alarmed by edges I might fall from, staircases I needed to negotiate, objects that floated before me and needed to be swatted or grasped with my glove. I could tell myself firmly that it wasn't real—the sketchiness of the landscape assured me of that—but my heart beat faster and I backed away when I saw dangerous edges or had to walk down imaginary staircases without banisters.

If you were observing, people in the goggles and gloves were clownish, stepping high over imaginary obstacles or batting imaginary floating objects. After you'd donned the helmet and glove, it looked real enough. But only enough: the landscape was sufficiently suggestive for you to behave as if it were real, although you knew it wasn't. VR would come to have applications in medicine, military training, PTSD treatment, and entertainment. Recently at the Jewish Museum in New York City, I visited an architectural exhibit that allowed viewers to see, through VR glasses, an architect's furniture designs in imagined room settings.

Lanier and I immediately hit it off. We cheerfully argued philosophy with each other while Lanier's businesspeople fumed—after all, real, not virtual, customers were waiting in the reception room. They needn't have worried. Sitting with us, waiting to take his turn, was

Alex Singer, a Hollywood producer, whom we also knew from an informal business network we belonged to. Singer would provide plenty of business for VPL and later, for Lanier as a consultant.

After VPL was disbanded (Lanier recounts its founding and complicated ending in his 2017 memoir, *Dawn of the New Everything*), Lanier left Silicon Valley and came to live in New York City, where Joe and I got to know him better. At what must have been tremendous expense and trouble, he brought along just a few of the unusual musical instruments he collects; they adorned his Tribeca loft like elegant sculptures. If I was lucky, he'd float among them and play a few, usually so exotic I'd never before heard the sounds they made. He knew their names, their histories, and their connections with similar instruments all over the world. One day I was in the Metropolitan Museum and stopped in the musical instrument collection. For a moment, I longed so strongly for Lanier to appear and explain some of them to me that I wasn't even surprised when in fact he did appear—flowing dreadlocks, in white from head to toe, but with a friend, and apologetically, too busy to linger.

Lanier's music is as fundamental to him as any technological skills he has. He often invited us to The Kitchen, an experimental New York City music venue where he regularly performed with other artists such as Philip Glass and Yoko Ono. One night we were lucky enough to join him for dinner with a young Sean Lennon, another musician he performed with.

The World Trade Center was very close to Lanier's loft, and after the 9/11 attack, he was prevented from going home for weeks. When he was finally permitted back, he packed up the exotic instruments and returned to California, to our regret.

Any evening with Lanier was—and is—always a treat. Ideas explode; some of them even stay aloft. In the days he was visiting at Columbia, he was helping produce a kind of virtual reality system for heart surgeons that would eventually allow surgery without breaking the breast bone, a project that has gone on to real success.

One summer day in Santa Fe, I knew Lanier was in town—we planned to be together the following day—so it didn't entirely surprise me to see him strolling toward me along Palace Avenue. I was with a friend and her very conventional visitor from Tulsa, Oklahoma. I stopped, gave Lanier a hug after not seeing him for a while, and we made last minute plans for the next day.

When Lanier and I parted, Ms. Smugly Conventional of Tulsa said: "Well! What a strange-looking individual!" After an entire lunchtime of such stuff, I'd had enough. "I believe your husband just had heart surgery," I snapped. "That strange-looking individual was probably responsible for your husband's successful outcome."

The following day, Lanier, Lena (who would later become his wife), Joe, and I drove to see another composer (Lanier not only plays every instrument under the sun, he also composes). This composer lived in a geodesic dome in the remote desert east of Santa Fe, an evocative journey for Lanier, who also once lived in a geodesic dome in the southern New Mexico desert he'd designed at about age thirteen, relying for its construction on a book he believed was sound, but was in reality only describing "ongoing experiments."

This day, we covered a long, bumpy trip over dirt roads, so Lanier entertained us by teaching us how to call goats, not as easy as you'd think. He'd herded goats to pay his college tuition after he'd skipped high school and gone straight from middle school to New Mexico State University in Las Cruces.

What was a young scientific genius like Lanier doing in southern New Mexico? His parents had both been immigrants, his mother a survivor of a concentration camp, his father an escapee from the pogroms of Ukraine. They'd eventually immigrated to the United States, where they met and married. Although Lanier was born in New York City, somehow the family found its way—fled?—to New Mexico, where his mother, trained as a dancer, supported the family as a kind of day trader. When Lanier was only nine, she died in an automobile accident. In *Dawn of the New Everything*, Lanier movingly describes his catatonic grief lasting for a year or so after her death. Father and son lived a nomadic life in tents, and then the geodesic dome. At age thirteen, Lanier persuaded New Mexico State in Las Cruces to allow him to take courses in science and music, and eventually he came to the attention of scientists like Marvin Minsky, who would later welcome him at MIT (Lanier, 2017).

Lanier has been consistent in his strong beliefs that computers and humans are not interchangeable in any significant way. We discuss it good-naturedly; in most respects, I agree. His earlier book, *Who Owns the Future?,* argues that you should be paid for any private information a corporation or government has about you in the same way someone who uses any property of yours would compensate you. He even suggests that this might be a way of providing at least a minimum income for those who'll inevitably be unemployed by technology (Lanier, 2013). Recently, he's elaborated on that pay-for-use-of-intellectual-property in terms of AI: improved algorithms improve themselves by learning from human accomplishments.

Don't those humans deserve some compensation for their contributions to smart algorithms? (Brockman, 2014).

With AI's present public prominence, Lanier has begun speaking out about AI itself. On the existential threat that some boldface names in the science and tech world have expressed about AI—for example, Elon Musk, Stephen Hawking, and Martin Rees—Lanier says that as much as he respects these scientists for their scientific accomplishments, he thinks they're placing a layer of mystification around technology that makes no realistic sense. If, on the other hand, their anxiety is a call for increased human agency—let's not allow bad things to happen with this new technology—then it serves a purpose. "The problem I see isn't so much with the particular techniques, which I find fascinating and useful, and am very positive about, and should be explored more and developed, but the mythology around them which is destructive."

This distiction between techniques and mythology is important. Of those layers of mythology, one of the most interesting is what Lanier sees as the confusion with religion, a magical, mystical thing. AI is not religion nor is it mystical: its abilities rest on the work of thousands, maybe millions, of human intelligences, which are being used without financial compensation. Translators, for example, become part of a victim population, as do recording musicians, or investigative journalists. Now AI becomes a structure that uses big data, but it uses big data. . .

> . . . in order not to pay large numbers of people who are contributing. . . .Big data systems are useful. There should be more and more of them. If that's going to mean more and more people not being paid for their actual contributions, then we have a problem. (Brockman, 2014)

Informal payoffs, as distinct from formal payoffs (royalties) are useless to people who actually have to pay the rent. With that I agree altogether, and I hope that the new explorations of ethics in AI will address this problem and find a fair, just, and ethical solution.

The mythology, Lanier believes, is a very old idea in a new costume:

> To my mind, the mythology surrounding AI is a re-creation of some of the traditional ideas about religion, but applied to the technical world. All the damages are essentially mirror images of old damages that religion has brought to science in the past. There's an anticipation of a threshold, an end of days. This thing we call artificial intelligence, or a new kind of personhood . . . if it were to come into existence it would soon gain all power, supreme power, and exceed people. The notion of this particular threshold—which is sometimes called the singularity, or super-intelligence, or all sorts of different terms in different periods—is similar to divinity. Not all ideas about divinity, but a certain kind of superstitious idea about divinity, that there's this entity that will run the world, that maybe you can pray to, maybe you can influence, but it runs the world and you should be in terrified awe of it. That particular idea has been dysfunctional in human history. It's dysfunctional now, in distorting our relationship to our technology. (Brockman, 2014)

And like many religions in the past, this mythology of AI exploits ordinary people in the service of the elite priesthood. Above all, it ignores human agency. We can shape our future legally and economically and in security.

As we'll see, others also believe we can and are working to make that happen.

These days, Lanier is settled in a house in the Berkeley hills with Lena and their young daughter, Lillybell, along with a selection from his musical instrument collection on three floors up and down the hillside. He writes and commutes to Silicon Valley. The house is witty: the last long conversation we had there was over delicious Russian tea while I sat in their living room on a four-poster Chinese bed draped in red silk. I suspect Joe and I regretted on his behalf far more than he did that, thirty years after VPL, Facebook paid $2 billion for a new virtual reality startup, Oculus VR.[1]

2.

Lotfi Zadeh was one of the best arguments I know for the tenure system. He arrived at the University of California, Berkeley in 1959, and because he'd already been tenured at Columbia in electrical engineering, he received immediate tenure at Berkeley. He'd been a brilliant young student in Tehran, his family's home. (However, he was born in Baku, Azerbaijan, where his Persian father was a foreign correspondent for an Iranian newspaper, and that city so captivated and influenced him that he wished to be buried there after his death, and was.) After college he made his way to the United States, where he received a masters from MIT and a PhD from Columbia. He taught at Columbia for ten years until he moved to Berkeley.

In 1965, academically secure, Zadeh published his first paper on fuzzy sets, a system, he'd claim, that allowed you to say something was "almost" there, or "not quite," or "very much" there. He once defined fuzzy logic as "a bridge between crisp, precise computer reasoning, and human reasoning" to me. It was a kind of approximate reasoning,

[1] VR was expected to transform video games, which it has. But as Corinne Iozzio notes in "Virtually Revolutionary," an article she wrote for the October 2014 issue of Scientific American, VR technology is also being used widely in psychological treatments for post-traumatic stress disorder, anxiety, phobias, and addiction and in aviation training. Later speculators imagine a dissolution between humans and the world, a kind of late-stage Buddhism achieved instantaneously with a headset.

which, Zadeh said, includes most everyday reasoning, such as where to park your car or when to place a telephone call.

Such problems can't be precisely analyzed because we lack the information for precise analysis. Moreover, standard logical systems, he argued, have limited expressive power. High precision entails high cost and low tractability, as if you had to park your car within plus or minus one one-thousandth of an inch. Fuzzy logic, on the contrary, exploits the tolerance for imprecision. Fuzzy logic, he said finally, was easy to understand because it was so close to human reasoning.

To say this idea was greeted with puzzlement by mathematicians and computer scientists is to put it generously. The theoreticians didn't know what to make of this strange logic, and if Zadeh felt he belonged among AI people, that camp was not merely puzzled but dismissive. Had Zadeh been a young assistant professor, following where his brilliant mind led, he'd eventually have been forced to follow a different career. With tenure, he was safe to stretch.

Joe and I met Lotfi Zadeh when we first came to Berkeley for one of our summer stays, and after we bought a Berkeley condominium, we saw the Zadehs often. Zadeh was slender, so spare that it seemed his flesh was barely sufficient to cover his skull, the cheekbones prominent, the forehead high and uncreased, large brown almond-shaped eyes that watched the world guardedly. He and his wife Fay were generous hosts, including us in dinner parties and their polyglot and musical New Year's Eve parties. I liked Fay very much. A striking, nearly life-sized oil portrait of her, clad in a sweeping pink evening gown, hung over a staircase in their Berkeley home, and yet it hardly did her justice. Fay had an enviable gift for languages, so her friends included women who spoke German, French, Japanese, Farsi, and any number of other tongues in which Fay was fluent. She was extremely warm, extremely practical. She was also kept busy working as Zadeh's personal secretary, because Berkeley was already on straitened budgets, and the amount of correspondence from around the world on fuzzy concepts was enormous—I'd see the stacks of letters on her desk when we visited.

Thanks not only to the success of fuzzy logic, but also to his diligent apostolizing (two-day round trips from Berkeley to Japan were commonplace), Zadeh was becoming famous around the world, if not yet in the United States. At dinner parties we usually met one or two of his foreign disciples, especially the Japanese, although followers came from everywhere. I wondered if fuzzy logic's attraction was something I've mentioned about expert systems, that it offered a spectrum of opportunities: problems for the brilliant and less brilliant to tackle. That's a formula for getting your ideas out in the world, something for everyone. As Zadeh himself had argued, fuzzy logic was close to human logic, so it was easy to understand. When I went to Japan and encountered fuzzy washing machines, fuzzy rice cookers, and fuzzy braking systems on the trains (because I didn't also know about fuzzy logic embedded in HDTV sets, camera focusing, or Sony palmtop computers) I commented on it to my Japanese hosts. They, in turn, were deeply impressed that I counted Lotfi Zadeh among my friends.

But the AI community in the United States still ostracized him. At a 1978 meeting of the International Joint Conferences on Artificial Intelligence in Boston, I stood in a hotel lobby among a group of AI people who'd been invited to dinner at a local professor's house. Zadeh passed by, hesitated for a moment, then saw no one was going to invite him, and continued on hurriedly. I felt an acute flush of shame and misery—it wasn't my party, so not my place to invite him, but the feeling was from that archetypal school birthday party, where some kids in the class are pointedly excluded.

Zadeh and I had lunch together one day in 1983. "Ah, my dear," he said philosophically. "It's a good news/bad news joke. The good news is that AI is working. The bad news is that it's fuzzy logic." He was right in an important sense: he was becoming not just famous, but acclaimed.

And appropriated. In London in June 1988, I came across an art exhibit at the Barbican of the newest, most promising young artists in France. One piece was called *Information:Fiction:Publicité:Fuzzy Set*. The artists, Jean-François Brun and Dominique Pasqualini, had mounted color photographs of clouds and sky in tall light boxes, hinting at fuzzy, hinting at whatever else. I sent the brochure to Zadeh to amuse him.

Zadeh was an ardent photographer himself, so no guest escaped the house without having a picture taken. Because the portraits on his wall were of famous people—Rudolf Nureyev in mid-*grand jeté*, Alexander Kerensky looking suitably melancholy at his life's outcome—I didn't mind holding that pose (whatever it was) every time I was with the Zadehs. Until he died, Joe had a picture of me in his office that Zadeh took in the 1970s, where I'm stretched out on the fine silk Persian rug in the Zadehs' living room, wearing bright blue slacks and turtleneck against the scarlet medallion of the rug. When I needed an author's picture for *Machines Who Think*, Zadeh took it in his Berkeley garden.

Over dinner one night, I heard this story from Tia Monosoff, who had been Zadeh's student as an undergraduate in the 1980s. She was taking the final exam in Zadeh's course in fuzzy logic, and for reasons she can't remember, had a complete meltdown and couldn't finish the exam. She walked out of the examination room and immediately called Zadeh, apologizing for and trying to explain her lapse. He began to question her about fuzzy logic, questions which now she could answer fluently. "Okay," he said after some minutes, "I'll give you a B." Expecting at best an Incomplete–or dreading worse—she was deeply grateful for his generous and wise understanding of how things can go awry.

In December 1991, I heard Zadeh give a lecture at Pasadena's Jet Propulsion Laboratories and realized then, with people standing, crowding the aisles, he'd become something of a legend. In that lecture, he showed viewgraphs of all the terrible things that had been said about fuzzy logic over the years. "Pamela McCorduck," he added, nodding at me, "has written enough about artificial intelligence to become something of a fixture in that community, and has probably heard even worse." (Actually, I hadn't. Nobody even took the trouble to think of insults.) Fuzzy, he continued, was still pejorative in the United States but had high status in Japan, such that there were "fuzzy chocolates" and "fuzzy toilet paper."

In short, Zadeh laughed at himself and won an already loving audience completely over to his side. "Funny Lotfi," I wrote in my journal that night. "Laughing last and laughing best."

Zadeh was always gracious and hospitable, but a membrane persisted between him and the world that was impossible to penetrate. The only time I came close was an evening when he called to invite me to dinner. I was alone and glad to see him. He too was alone. But although we two talked easily, he ate nothing. Why? I asked.

He said deprecatingly, “A little medical procedure tomorrow.” So I called him the next day to see how it had gone. He was ecstatic I’d remembered—“cool remote Lotfi,” I wrote in my journal, “so humanly pleased with an ordinary human gesture.”

By the mid-1990s, fuzzy logic had proved itself even to the doubters—and Zadeh had lived to triumph. He received the Allen Newell Award, which is presented “for career contributions that have breadth within computer science, or that bridge computer science and other disciplines.”[2] This award came almost literally from the same group that had once excluded him from that Boston dinner (and everything else). He was inducted into the Institute of Electrical and Electronics Engineers AI Hall of Fame, became a Fellow of the Association for the Advancement of Artificial Intelligence (as well as a fellow of many other distinguished professional societies) and could count twenty-four honorary degrees from all over the world.

Joe and I took the Zadehs to lunch in Berkeley when we visited a few years ago. They were each ninety, and although they showed their age in little ways, Fay was still scolding him as Lotfichen, and Lotfi was as intellectually sharp—and personally opaque—as ever. Fay was to die early in 2017, and Lotfi died on September 6, 2017 at the age of 96.

3.

In 1983, Gwen Bell, an innovative and gifted city planner, then married to computer architect Gordon Bell, adopted the modest corporate museum of the Digital Equipment Corporation and transformed it into The Computer Museum of Boston, relocated on

[2] The quotation is from the ACM SIGAI web page for the Allen Newell Award, retrieved from http://sigai.acm.org/awards/allen_newell.html

Boston's Museum Wharf. Its holdings would eventually form the kernel for Silicon Valley's Computer History Museum. In the early days, Gwen began a popular fundraiser for the museum, a trivia quiz called The Computer Bowl, which pitted two teams against each other, East and West, made up of people well-known in computing, to be broadcast on a nationally syndicated television show, *Computer Chronicles*. When Bill Gates, then the CEO of Microsoft, participated in an early episode as a member of the West team (and earned MVP status) he was hooked, and for the rest of its nearly ten-year run, he was the quizmaster.

Gwen invited me to be captain of the East team in 1991. "You didn't!" Ed Feigenbaum groaned to me on the phone. "What if you're humiliated…?" I hadn't thought of that. Entirely possible. Pie in the face, dunked into the tank: hey, it was a fundraiser.

The Computer Museum chose superb teammates for me: James Clark, vice president for high-performance systems of AT&T (and African American, unusual in the field); John Markoff, who then covered the computer industry for *The New York Times* and had written books about it; John Armstrong, vice president for science and technology of IBM; and Sam Fuller, research vice president of Digital Equipment Corporation (Nichols, 1991).

Because everyone would expect us to appear all uptight Eastern, tie and jacket, I suggested to my teammates we do the contrary. Maybe they'd wear whatever outfits they used for exercise? John Markoff rolled his eyes: I am *not* going on TV in bicycle shorts!

But we all got into the spirit. One team member surprised us with black satin team windbreakers, and under these we wore sassy tee shirts and easy pants. We each had on baseball hats (backward, of course, fashion forward at the time). John Armstrong's was the John Deere hat he wore when he mowed his lawn; James Clark wore a Top Gun lid. I'd asked a ten-year-old skateboarder to go shopping with me and found an oversize Daffy Duck t-shirt, skateboarding pants. I also borrowed my west coast nephew's skateboard, plus his hat, which read: "If you can't run with the big dogs, stay on the porch." Decked out so, the audience was laughing the moment we marched to our buzzers. It gave us a quick psychological edge over the business attired, and openly astonished, West team.

I'd seen Bill Gates arrive earlier at the San Jose Convention Center, where the quiz was televised, and was surprised that he had only one assistant with him. I imagined that one of the richest men in the world would be surrounded by a phalanx of bodyguards and gofers, but no. This was a hang-loose Bill Gates who was very endearing.

The professional host, Stewart Cheifet, moved things along; Bill Gates asked the questions; the buzzers sounded; and the East team, led by their skateboarder captain, won handily, 460 to 170 (Nichols, 1991). I won MVP status, something to share with Bill Gates. Maybe our goofy outfits had relaxed us or maybe it was just our night. For sure we were lucky to have well-distributed arcane knowledge. I watched myself later, amazed at how cool, even snooty, I seemed. In fact I was determined not to be humiliated, as Ed Feigenbaum had warned, so I was nervous and concentrating hard.

"A computer historian!" Dave Liddle of the West team protested. "No fair! Of course they were gonna win!"

4.

In May 1986, an editor at Harper and Row, Harriet Rubin, surprised me. How did I feel about collaborations? My only collaboration up to then, *The Fifth Generation*, had been great fun. Moreover, I was surrounded by scientists who loved to collaborate, which not only amplified each individual's work, but also made it less lonely. With the right coauthor, I was open. She told me John Sculley, who'd been brought in by Steve Jobs to run Apple and then fired Jobs, was looking to do another book like Lee Iacocca's best-selling *Iacocca: An Autobiography*.

"Ah," I said, "he wants a ghostwriter. I'm not sure I'm right for that." No, he wanted a collaborator, was even "willing to share the credit." I didn't say no at once, but I was dubious. Still, it would be interesting to meet him. It might even be a provocative project. "There's writing for the movies, and then there's writing for the movies," I wrote in my journal.

I reviewed Lee Iacocca's as-told-to memoir. Iacocca had been the anointed successor to Henry Ford II at Ford Motors, but capriciously, or feeling threatened, Ford had suddenly fired his crown prince. Iacocca was stunned, deeply hurt, but took the best possible revenge: he went to Chrysler, which was nearly bankrupt, and turned it around, and Chryslers began to outsell Fords (Iacocca & Novak, 1984).

Iacocca had two splendid myths going for him—first, immigrant rags-to-riches (though his father had been the immigrant); and second, the crown prince exiled by the old king, who finds another kingdom to rule, with subsequent success even grander than the old king's. But the underlying myth in Sculley's conflict with Steve Jobs sounded like the old king slaying the crown prince: an idolized victim exiled, effectively slain, by a bean counter. At best, it was Cain and Abel. At this time, Sculley had not yet saved Apple, and Jobs had not yet found a kingdom where he could outdo Apple. It would be an enormous writerly challenge, but not impossible, I thought, if I put my mind to it.

Everyone seemed in a great rush. They'd want a manuscript at the end of the summer, and it was already mid-May. Another writer was doing Steve Jobs's side of the story, so Sculley's and my book would be a riposte, or better, a preemptive strike. That was possibly more interesting: at least it could aspire to human drama, and not be just another forgettable piece of businessman's lore. Sculley was seeking someone who was strong-minded, who insisted her name be on the spines of books. If he'd wanted merely a journalist or a ghostwriter, thousands more pliant were available.

Jane Anderson, a young Englishwoman who was Sculley's personal PR person, invited me to lunch in San Francisco. The Rosenkavalier, she was very open that they wanted me and asked what would move me? That it be more than just another businessman's book, I answered, and quoted Melville on mighty books, mighty themes. That seemed to please her. I wasn't antibusiness; I thought great literature could come from anywhere, approached with suitable intelligence, complexity, and freshness. Sculley was very shy, very private, she said; had begun in design—which surprised me—and had an avocation in science and technology. She gave me some further background, some of Sculley's thoughts, some of his memos, and suggested I call Alan Kay for a reading on Sculley. Kay was very positive: a man "who loves ideas;" I'd enjoy working with him.

But the memos Anderson gave me were uninspired, and I still didn't see any way of dealing with the underlying myth problem, except to portray Sculley as a rounded, vulnerable, contradictory human being, which I was sure no CEO in his right mind would permit. Early in his career, Sculley had turned from design to marketing because he saw that in corporate America, marketing made most of the decisions. Yet marketing values were shallow, soda pop values. Was this why Sculley could be lured to Apple, to return to meaningful values?

I was still mulling all this when John Sculley, Jane Anderson, and I finally had dinner together. Shy he might be; egoless he wasn't. He wanted the book to be told in the first person (so by a ghostwriter after all) and wanted final say on every word. Although Kay had said Sculley loved ideas, I waited for a surprising idea from him, but none came. Perhaps he was saving them for the book. So I probed. Why a book, since most of us write for fame and fortune, and he had ample amounts of each? "I think I have a book in me," he said, with a sweet ingenuousness that made me smile.

Aloud I mused on the underlying myths of the Iacocca book, a book he admired: rags-to-riches immigrant, old king banishes threatening crown prince. Then I examined the underlying myths of his own story, interloper banishes the crown prince, at best, Cain and Abel, which startled him. It isn't insuperable, I told him. But it will be difficult, and take some imagination to solve the problem. It couldn't be just puffery, or people would ask why he'd got me to write it when he already had first-rate Silicon Valley PR people.

But I could see I'd lost him. Bean counter banishes crown prince? Cain and Abel? Who wants to be part of those tales?

Jobs was difficult—how difficult we wouldn't really know until Walter Isaacson's biography, *Steve Jobs*, was published after Jobs's death. An impossible situation had developed between Sculley and

Jobs at Apple in the mid-1980s, and at the time, Apple's board of directors backed Sculley. Jobs had to go. Sculley would indeed preside over a period of great profitability in the late 1980s at Apple, although die-hard Jobs supporters argue that Sculley was cashing in on new products Jobs had already put into place. In the turning of the corporate wheel, Sculley himself was eventually ousted from Apple, and Jobs came back and made Apple into an even more profitable company, with products that were globally admired and emulated. Sculley went away to be an extremely successful entrepreneur, investor, and businessman.

I liked Sculley personally, and when I ran into him at Brown University two years later, where, as an alumnus, he was much involved with a new computer science building, I reintroduced myself and said I hoped we might find something to collaborate on. I'd just begun work on a book about art and artificial intelligence, and he said politely that he was looking forward to reading it.

Meanwhile, I'd met Steve Jobs, the exiled prince, at a cocktail party to celebrate the inauguration of his NeXT machine, which he hoped would indeed be the next big thing, a way of reclaiming his kingdom. To my embarrassment, Joe told Jobs the Sculley story. When he heard Joe repeat my phrase: "Iacocca was the prince who, in exile, bested the old king, but you banished the prince," Jobs suddenly stopped being the gee-whiz kid. Joe brought him over to me for corroboration. Yes, I said, I'd presented this to Sculley as a problem for any writer who undertook to tell the story. But for Jobs, of course, it was his life. He grasped my hand. "You really said that to him?" he asked with great intensity. "Yes, of course." Jobs's young face was struggling with many emotions. He was nearly in tears. He didn't let go of my hand. “Thank you,” he said gratefully. “Thank you for telling me that.”

5.

I sometimes wonder how the AI pioneers would regard present-day Silicon Valley. They’d be very pleased that AI is so prominent, highly honored, and pursued. They might perhaps be amazed that by mid-2018, the FAANG group of firms (Facebook, Amazon, Apple, Netflix and Google, AI firms all) was worth more than the whole of the FTSE 100, according to *The Economist.* They might be less enchanted by a culture that revolves so single-mindedly around making money. Each of AI’s four founding fathers lived modestly, in houses they’d acquired when they were new associate professors, houses where they’d brought their children up, where they ended their days. Science, not the acquisition of capital, drove them. Each of the four had strong if varying senses of social justice, and would be troubled by how much of machine learning learns, draws conclusions from, and reinforces unexamined social bigotries. That the social spirit of Silicon Valley mirrors the most retrograde of other commercial sectors, finance, would dismay them. They would have wanted something better, more honorable, of their brainchild.[3]

[3] And oh, the problems of casual, inept, cut-rate, or overweening applications of machine learning. Virginia Eubanks’s Automating Inequality (St. Martin’s Press, 2018) is a horror story of rigid and brittle systems ruling over and punishing the American poor. Andrew Smith’s article for The Guardian, “Franken-algorithms: The deadly consequences of unpredictable code”(August 30, 2018), deserves a book. Yasmin Anwar’s Berkeley News article “Everything big data claims to know about you could be wrong” (June 18, 2018), describes the follies of averaging over large groups in say, medical outcomes. Clare Garvie’s story for The Washington Post, “Facial recognition threatens our fundamental rights” (July 19, 2018), speaks for itself. This threat is already operational in China, with 1 in 3 billion accuracy of face recognition, which monitors citizen behavior at a level where individuals earn “good citizenship points” for behaving exactly as the state wishes, and demerits when that behavior is considered bad. On the other hand, Thomas McMullan’s story for Medium,“Fighting AI Surveillance

It would be left to me and many other women to be impatient, and then angry, with the sexism that dominates Silicon Valley. But that moment was to come.

Meanwhile, AI had moved into that empty mansion of the humanities and wasn't just cleaning it up and straightening things out, but was making major alterations, as we'll see in the next part.

with Scarves and Face Paint" (June 13, 2018), shows guerillas are now inventing electronic scarves and face paint. In The Washington Post story "Microsoft calls for regulation of facial recognition" (July 13, 2018), Drew Harwell notes that Microsoft has officially called for government regulation of facial recognition software, as "too important and potentially dangerous for tech giants to police themselves."

Part Six: Arts and Letters

A story travels in one direction only,
no matter how often
it tries to turn north, south, east, west, back.

—Jane Hirshfield, "Tolstoy and the Spider"

Art and Artificial Intelligence

1.

Painter Harold Cohen's work thrust him into the center of one of the 20th century's most contentious conflicts—it endures yet—the war of authenticity. You've heard it before. "Is it really thinking?" For him, the question is also "Is it really art?" In time, there'd be guerilla actions around creativity, learning, the new role of the artist, and the appropriate role of the computer. Writing *Aaron's Code* (1990), a book about Cohen and the ways he used AI to create art, would bring me face to face with the same problems I'd met writing *Machines Who Think.*

Cohen's work fits into the traditions of Western art in two major ways: The first is self-portraiture. A long tradition, reaching back at least to the early Renaissance, has honored artists who offer deep and provocative self-portraits. The difference in Cohen's work is that the self-portrait is dynamic (that is, it changes over time) and it's a portrait not of the artist's physiognomy, but of his cognitive processes as he works. The essential work of art, one might argue, is the program called Aaron, not necessarily the images that Aaron produces—though they are the physical evidence that code has captured cognitive processes to a significant degree.

Self-portraits allow us to imagine that we can detect the artist's emotional state, not his cognitive state. Contemporary psychology gently corrects us: the cognitive and affective cannot reliably be separated. In any case, surely Aaron's actual code is the result of a consuming passion: from its first lines of code, Cohen spent more time with Aaron than with any human being, and that accounting held for the rest of his life.

A self-portrait that captures the artist's cognitive processes to a significant degree and in a dynamic fashion is surely a new thing under the artistic sun, which allows Cohen another major place in Western art, namely as the begetter of profound, even revolutionary, innovation.

Philosopher Alva Noë (2015) argues that our lives are structured by organization. Art is a practice for bringing our organization into view; in doing this, art reorganizes us.[1] If so, Cohen's work fits the grand artistic tradition this way, too.

What Cohen accomplished seems very difficult for most of the art world to grasp. Since the publication of *Aaron's Code* in 1990, digitally manipulated images have become more familiar and have been admitted in some degree to the canon. Art produced by machine learning has also created a modest stir. In October 2018, a machine learning–generated image printed on canvas, called *Edmond de Belamy from La Famille de Belamy* and created by a Parisian group,

[1] Noë further argues that "technologies are organized ways of doing things. But this equivalence has a startling upshot, one that no one has noticed before. Technologies carry a deep cognitive load. Technologies enable us to do things we couldn't do without them—fly, work in a modern office place—but they also enable us to think thoughts and understand ideas that that we couldn't think or understand without them." In that sense, AI is a technology as well as a science.

sold for $432,500 at Christie's (Cohn, 2018). But the depth of Cohen's achievement is still unfathomable to most curators and collectors.

In the 1960s, Cohen's reputation as a painter in his native London was soaring. By 1966, he was one of five artists who represented Great Britain in the 33rd Venice Biennale, and his work could be seen in important galleries in England and the Continent. Although he played a central role in the London art world of that era, 1968 found him restless, ready for some kind of major change. That fall, he arrived in San Diego, California, with three young children (his first marriage had ended, and he retained custody). He settled down to paint and teach in the newly established visual arts department at the barely decade-old University of California, San Diego, beautifully situated on the coast just north of San Diego in La Jolla.

Cohen was a stocky man of medium height, with a rich rabbinical black beard and graying hair pulled back in a ponytail. Behind his glasses, his dark intelligent eyes seemed portals to an unusually complicated soul. Without hesitation, he could speak on nearly any topic, his language impressively Mayfair (unless he lost his temper, when it slipped into the East End, where he'd grown up). He was also sharp-tongued and dismissive of many of his fellow artists, although he once said to me: "I value less and less in art these days, but what I do value, I value deeply." He meant Cezanne; he meant Duchamp.

Jef Raskin, later to have a hand in designing the first Apple Macintosh, was a colleague on the visual arts faculty at San Diego. Early in Cohen's stay, Raskin said almost truculently: I can teach even *you* how to program. Cohen took it on, thinking it might be as interesting as doing crossword puzzles, one way he passed the time as he mulled a painting.

Cohen had first seen computers in action in a 1968 London show called Cybernetic Serendipity. It was the heyday of "computer art," when anything that could be digitized, processed, and printed with a plotter ended up on gallery walls. Either computers were very stupid, or people were doing very stupid things with them, he thought.

But by learning to program, he slowly (and in his recollection, independently) arrived at the same insight that AI researchers had from the beginning: the computer is a general-purpose manipulator of symbols, and thus can be viewed as *functionally* equivalent to the brain.

2.

Cohen conjectured that AI might be a means to test some of his theories about making art. With a program, he could model a theory, watch the output, and then revise the program (or the theory) until the output was right. What did right mean? He believed it to mean the *evocation,* not the *communication*, of meaning between the image and viewer. Art was a meaning generator, not a meaning communicator.

With the program called Aaron (his own Hebrew name) Cohen was beginning to externalize knowledge that, until then, he'd held internally, often unconsciously.[2] Aaron knew and followed some general rules about making art on a two-dimensional surface. For example, the program knew how to represent occlusion (one object hidden behind another); how smaller objects at the top of the picture plane appeared to the human eye to be back beyond objects in the foreground. Aaron decided where to begin a drawing, which shapes

[2] Recall early expert systems, where a knowledge engineer evoked knowledge from the heads of experts and turned it into executable computer code.

and how many of them to include, and decided when it was finished. Once a drawing was begun, human intervention was forbidden. Owing to chance elements in the process, each drawing was different from any other; each drawing was an original.

Aaron was autonomous, not in the trivial sense that it could control the movements of a pen, but in the sense that it could invent those movements. It generated images instead of merely transforming them. For Cohen, then, the computer was another artist's tool, but of a different order from ordinary tools.

3.

Human artmaking is a fluent set of decisions based on the artist's awareness of the work in progress. A program to model that behavior needed a similar awareness. But in those days computers had no eyes to gaze at a work underway. Cohen wrestled with that problem in various ways, not as a psychologist proposing a model of human perceptual mechanisms, but as an artist, trying to fashion a model of art-making that would prove its plausibility by—what else?—making art. As Alex Estorick (2017) puts it, "Aaron had to learn to see in the dark." If it had no eyes to see, Cohen would give it the functional equivalent of eyes, an imagination so powerful it could envision a drawing, constantly referring to the drawing's totality in order to make the next mark on it.

What emerged was an arrangement of nested Russian dolls, Chinese boxes: a hierarchy of levels of conception. At the highest level was the human artist, Harold Cohen, who'd conceived the whole scheme, benignly hovering over the next conceptual level, his computer program, Aaron. Aaron was an entity with some general knowledge about artmaking and the capacity to make artifacts based on that knowledge. Finally at the bottom of the hierarchy (although, paradoxically, always the most visible feature) were the drawings themselves, each unique, unseen before, and not to be repeated. Cohen had vaulted to the plane of meta-artist, having created a work of art—the program Aaron—that itself made art. This was conceptual art of an unprecedented degree: for sheer nerve, Cohen was the equal of his spiritual forebear, Daedalus. Over the years, Aaron would grow to some 14,000 lines of code and be recast in different programming languages.

In my early AI days, Cohen and I often ran into each other at AI conferences, the only nontechies there, though Cohen's technical knowledge far exceeded mine, and he picked the brains of the AI people cheerfully to help him write his art-making program. By the early 1980s, Aaron was already making abstract drawings of recognized aesthetic value. The artist was unquestionably Aaron—it had learned how to draw from Harold Cohen and drew all the time.

4.

With all its art-making knowledge, Aaron was a kind of expert system, but also what Cohen called "an expert's system," the instantiation of everything Cohen knew about art and knew how to tell the computer. The program was becoming a singular expression of the artistic processes of a particular artist's mind, laid out in executable computer code. Aaron was contingent. It followed general rules, but even knowing those rules, an observer couldn't predict what the program would do: it moved through such a rich decision tree in the course of making a drawing that, again, no two were ever alike.

In 1983, Harold Cohen was invited to mount a show at the Brooklyn

Museum, where Aaron's drawings were exhibited, and viewers could watch the program make drawings in real time. Aaron's work was abstract then, with primitives like angles, combs, closed forms, and so on. Part of the excitement about Aaron was that it was a computer program, something just coming to public attention with the popularization of personal computers. This one was making drawings! Most viewers hardly grasped the intellectual claims Aaron could make—or would've believed them.

Joe and I went to see that Brooklyn show, thronged with curious viewers, and bought a couple of hand-colored drawings. At the time, Aaron could not color, and Cohen doubted it ever could. (Thirty or so years later, he solved that problem sumptuously.) We invited Cohen home for supper. He was inspiringly articulate about what he was up to, and it was a pleasure to see a *New York Times* art critic, Grace Glueck, take Aaron seriously and write a sensitive review of the Brooklyn show.

5.

Three years after the Brooklyn show, when I was writing a book about expert systems with Ed Feigenbaum and Penny Nii, Cohen suggested I should next write a book about him and his work.

September 30, 1986:

> *Harold here for dinner tonight, and I surprised myself a little by saying yes to doing a book about him. But his ideas are fascinating to me, and I don't think the effort will be great, considering the payoff: my high road to learning all about art.*

Unfortunately, by the time I began research for the Cohen book, both the artist and I were in trying circumstances. Cohen's second marriage had broken up, distressing him deeply. Joe and I had moved from New York City to Princeton, New Jersey, where Joe joined the computer science faculty at the university, but his main job was to run one of the National Science Foundation–sponsored supercomputer centers.

I conceded in my journal that I'd had six grand years in New York, and it was Joe's turn to do what he wanted. But Princeton was difficult—my life, social and professional, was in New York: I was perpetually taking the hour and a half train ride to the city.

February 17, 1987:

> *The Cohen project fills my mind. I think Harold has brought me back to my own art. In a sense, I'm using him to learn from. He has truly, importantly—and in a less important but literal sense—taken art where it has never before been.*

February 23, 1987:

> *Ed tells me Ray Kurzweil has made a film with a segment about Harold, gorgeous to look at, but neglecting to mention that the colors were supplied by the gifted hand of Harold Cohen. Ed said this publicly after the film was shown. Kurzweil's deputy went into earnest conversation with Harold, which Harold later told Ed amounted to: how can we get Feigenbaum to shut up?*

February 28, 1987:

> *Re-reading* Telling Lives [an anthology of work by biographers on the art of biography] *I have a sudden insight as to why I couldn't do the Simon biography. Right at the beginning Herb laid down a rule: nothing personal. This was, do not mention my family. I agreed, thinking it could be a book of ideas. But suddenly, ten years later, as I face the problem again, I realize Herb cut away from me what I not only knew how to do best, but also a vital part of*

> *the life. That limitation made the task impossible in any real sense. Odd that I never recognized this until now, and publicly and privately blamed myself alone.*

Nervously I gave a presentation on Cohen's work at John Brockman's New York City Reality Club on March 5, 1987. Afterwards, I wrote in my journal:

> *As it turned out, the Reality Club presentation was fun, though my own agenda was pushed aside in the uproar over IS COHEN DOING ART? To my astonishment, Joe and Freeman showed up. (John Brockman on the phone this morning: "Who else but McCorduck would have her own private claque consisting of Joe Traub and Freeman Dyson?") We'd met Freeman on a walk in Princeton a week earlier, and discussed the Reality Club, me saying later to Joe, I hope he doesn't come and I don't want you there either. Red rag: Joe cannot resist. Well, they did rough me up, but I gave as good as I got, and found myself enjoying it to a high degree. When John called this morning, it was to say they'd voted me best presentation of the year—an exaggeration, no doubt, but sweet to hear. Dumbfounded to see Benoît Mandelbrot there, but he behaved himself nicely; Hugh Downs next to me, scribbling notes furiously, though probably not for his TV show. Don Straus told me he'd forsaken I. I. Rabi for my talk—uh-oh, I thought.*

November 8, 1987:

> *Heard Larry Smarr give a marvelous talk at the University of Illinois. He's hired several artists, among them Donna Cox, to turn the rush of info from the Illinois supercomputer into visually accessible forms. Just wonderful, though curmudgeon Harold isn't impressed. Smarr's group is exciting, and whether their work is art, heaven knows it's important science. Harold argues art isn't in the service of science, but the artists feel, I think, they're getting a fair return by having access to the supercomputer. The images all that number-crunching produces are theirs to carry forward, and they do. Meanwhile, they're*

> *permitting scientists to see things never before seen—the collision of supernovas, for example. Great stuff.*

Joe and I went to London for Christmas that year and were able to see some of Cohen's work at the Tate. A 1963 painting he'd just sold to them, *Before the Event,* seemed to be doing then what was more than twenty years later suddenly so fashionable in New York art circles, quoting ideas and icons from science and transmuting them; in this case, replication—signaled by the central image, which was to my eyes, the primal copulation, surrounded by DNA chains, and what looked to Joe like state space diagrams. Ribbony images foreshadowed Aaron (unsurprisingly) and the bold glorious colors were unmistakably Cohen.

6.

Writing about Cohen's work wouldn't be easy. I had to educate myself well beyond my college art survey course and the naïve pleasure I took in museums and galleries. Work enough. But I also had to learn exactly what Cohen was doing. At the time, Aaron had turned from abstract to representational art, something the human artist never did. Each picture contained people, shrubs, trees, flowers, and rocks, although how many of each, what kind of each, and where they were placed, Aaron decided as it went along.

Cohen himself seemed moody and often unreachable, in great despair over the breakup of his marriage, over his advancing age, over his lack of recognition for this breakthrough effort, over any number of things. Thanks to the Princeton move, I was hardly my serene self. On March 30, 1988, I wrote in my journal:

> *The worst moment is when John Brockman yells at me for even considering doing the Cohen book. His reasoning: publishers want books that "jump off the*

> *shelves," Cohen is unknown in NYC and the art world, so only nerds would be interested, and nerds don't buy art books. It'll be poison for my future, since I'll go from being an author who makes money for publishers to an author who doesn't…yelling all my worst fears, full volume in my ear. I hold my ground, countering that Cohen is ahead of his time, a place I've also been; that this is to bring attention to Cohen in the art world (if that matters so much); that my life isn't dedicated to making money for publishers. Most of all, I need desperately to grapple with ideas again.John is no philistine and has pushed more than his share of cutting-edge ideas in the face of establishment skepticism, even scorn. He admitted later he was only doing his job as an agent. Push it as far as it can go, make it big and important, and it'll work, he said finally. Which answers the question of whether I focus narrowly or widely. But I was really down. The idea of doing another* Machines Who Think*—trying to convince editors that the topic is important—shrivels me.*

Cohen would swing through the New York area from time to time, and on a ramble through the Institute for Advanced Study woods in Princeton, we agreed that the book should embrace the history of ideas, as wide-ranging as possible. I didn't tell him what Brockman had said.

In mid-May 1988, Joe and I were having dinner with artist Lillian Schwartz and her physician husband, Jack, and got to musing about why computer art was so relatively stagnant. Schwartz agreed. "It's the software packages," she said finally. "They give easy access to artists, but not mastery of the medium. So most think they should go on doing what they're already doing, only faster and easier, and they're surprised it isn't altogether like that. Moreover, they don't imagine doing new things, locked in as they are to doing the old things 'faster and easier.'"

A year or so later, when I saw Lillian Schwartz in Utrecht, Netherlands, at an electronic arts conference, she added this insight: the blank canvas presents a fierce challenge to overcome, whereas the computer always has an easy way of beginning: a menu, a mouse, a program that begins and prompts your participation. So not only is the initial challenge lessened, but the continuing process is eased.

Harold Cohen spoke at that same conference, saying user-friendliness is an alienation from the tool. He charged that, by using packaged programs instead of writing their own, artists were evading the use of their own tools. Later, Harold added privately that "us old guys" already knew how to make art before the computer came along, but for youngsters who were just feeling their way, the machine overwhelmed them before they had a chance to find out what art is.

Maybe. Word processing offered some of this same ease to writers, but I didn't notice the essential part of writing was therefore easier. Of course I was one of the "old guys:" I'd learned to write with pen and paper, a typewriter, carbon paper, erasers; only midcareer with a computer.

7.

In June 1988, Joe and I went back to London, where we met Timothy Cohen, one of Harold's sons, an artisanal jewelry-maker. I wrote in my journal:

> *He arrives all dark, handsomely Byronic, with what turns out to be an incisive mind, willing to talk about his father's work in loving and perceptive detail: the fallowness of the early California years, the necessity of relating the earlier paintings of the Sixties to the work now. Thinks a color machine will be a disaster for Harold financially, in the sense that it will mass-produce the last hand-done thing, an event the art world wouldn't countenance—the rich will*

do everything to protect their investments. I agreed, but if you're on the correct side of history, then all that is a rearguard action.

Timothy Cohen was talking about art as positional goods, a term economists use for objects that are valuable not because they're one-of-a-kind or inimitable, but largely because other people can't have them. The art world had been about positional goods for a long time. Aaron, in its sly way, exposed this yet again.

We talked about technology changing the way art is done—oils permitted painting on canvas, which, hand in hand with other historical forces, brought about humanism. The question is what computing will bring about with art. I said I honestly didn't know.

Joe and I went on to Paris, where we brought the topic up with friends over long Parisian dinners.

What pushes an artist out of doing the usual very well, and into doing the new, the difficult, sometimes revolutionary? Yes, our culture is a bit odd in valuing the new the way we do—you could scarcely imagine conducting a puberty ceremony in a non-Western culture with a whole new take on the masks, say; and the Chinese valued sticking to the old forms. Economic issues: Could Aaron be pirated? Timothy worried about the glut of Aaron drawings: people wanted a signed drawing, not just a drawing. But that could easily be faked—Harold's changeable signature, or a specific one for Aaron. Then the collectibles might be "early Aaron," "middle Aaron," etc. And suppose Harold could endow Aaron with more intelligence than it has now, and it began to develop autonomously, even posthumously? Would each version of Aaron develop differently, given a few statistical differences in the actual employment of the program?

A posthumous Aaron would have its own problems. We don't desire an eternal late-Verdi-opera composing machine, or a few more Otello*-like operas. If we*

> *want operas at all, we want those that seem to connect with issues and styles that are now. So art is a conversation among the so-called human verities (themselves ever subject to change), the Zeitgeist, and the expression of an individual artist—all three are necessary. Finally, so much is chance. If you're lucky, like Bach or Donne, some Mendelssohn or T. S. Eliot exhumes you and champions your work. Or, you stay more or less continuously valuable, as Beethoven and Rembrandt have. Or, you enjoy a flurry of posthumous fame, and then disappear. All very capricious.*

July 25, 1988:

> *Saturday night to dinner at Cathleen and Peter Schwartz's, where his business partner, Jay Ogilvie, brings Doris Saatchi. We muse on why for the most part computer art hasn't moved on since the '60s. Doris, deep in the art world, has several conjectures: that no theory has developed…that much of the market [art buyers] is fundamentally nouveau, uncertain of its tastes, and like the 19th-century Pittsburgh nabobs who built replicas of known architectural masterpieces, the new buyers want the conventional paint-on-canvas, preferably certified by this "new mid-life-crisis career of the wealthy, especially women, called art consultants. Art by the yard." Also the problems of the poor materials contemporary artists use. She uncrated an Anselm Kiefer and the pile of sand at the bottom of the crate was so large the cat headed straight for it. Dishes keep falling off her Julian Schnabels. What do you do? I asked. Glue them back on, she said.*

8.

Aaron raised questions about originality, authenticity, intelligence, the meaning of art, its evaluation, but I began to think of it as also within another of the great traditions of Western art, the representation of knowledge—in this case, the representation of what Harold Cohen knew about artmaking. But Aaron went well beyond that.

Along with the stimulating questions, difficulties arose. What Cohen was telling me in our long interviews ("the tale Harold has created for himself" I called it) was orderly and rational, fair and high-minded, but it also suggested the well rehearsed (no sin, necessarily) and eventually raised more questions than it answered. Over dinner one night I questioned that smoothness. He agreed; felt he was gliding over the same material. He was extremely self-protective, I said, even evasive. "You want to fade out of the book entirely, but that will turn it into a PhD dissertation." Becky Cohen, his estranged wife, had used a simile: ideas are like parasites, they need a host.

After a few days of testiness between us, the artist said he was ready to try harder. It was, he agreed, one part Brit stiff upper lip, one part not answering the implicit question, only the explicit. "You must ask: wasn't the isolation awful? And I'll say yes, I hadn't remembered, but it was. I'd cut myself off from everything, and at one point thought I'd gambled my entire career and lost. There were years when nothing seemed to be happening: UCSD thought it had hired a big-time painter, when all they got was somebody who'd disappeared into computing." Becky Cohen had compared it to Jacob wrestling with the angel in the desert, and typical of Cohen: very private, nobody really knew. Except it went on for twelve years.

And what was I trying to do here? Harold was offended by drafts I sent him and couldn't understand why I'd detected not only paternalism in his relationship to Aaron, but a firm streak of misogyny, which I thought figured into the art. (When the book was finished, Becky Cohen wrote me: "Yes, yes, how did you know? He repelled two wives and a daughter with it!")

On August 19, 1989, I wrote a letter to Cohen, recorded in my journal:

> *I aim at grasping the life and the art as a series of intertwined, mutually nourishing patterns. My job is to find those patterns, particularly when they wouldn't be apparent, and illuminate them, pointing out how the life informs the art, the art informs the life. The task doesn't involve censure, it doesn't involve much praise (though this I lapse into from time to time; can't suppress it). It involves delineation and explication. Period.*

First Cohen's estranged wife Becky Cohen, and then Harold himself had asked me why I wasn't consulting other experts. I replied in the letter:

> *The only mechanism I've had confidence in is my own observations, coupled with my own interpretations. I assembled the data. I tried very hard to understand it from your point of view; I studied the discrepancies I saw between your point of view and mine. I've stepped back again and again to understand it all against the larger culture of which we're both a part. I have confidence in such a way of working because that's how I wrote MWT. A casual reader might think I used all those interviews in MWT to check and counter-check. In fact, nothing of the sort. Everybody had his own version of the story, some more intelligent than others, but none of them was particularly satisfactory alone. So I did it myself. In other words, the aggregate of interviews for that book played the same role as my many interviews with one person here for this book. In the end I have to trust my own intelligence.*

And then I added by hand: "And be prepared to fail."

The letter went on:

> *Meanwhile as I very self-consciously understand it, I am busy fashioning a linguistic construct of your art and life myself. If the maker's hand is apparent, I am doing it as honestly and dispassionately as I know how. The dispassion*

> *doesn't entirely preclude partiality; I couldn't imagine spending two or more years of my life on a subject I didn't really admire: I admire it/you, you know that. I say it once again in case it got by you. It's a different personality that spends its life on a topic it ultimately wants to trash (though such biographers exist—curious). That the book isn't unalloyed valentine—well, my gift is to love profoundly, not blindly.*

I was glad to put it all into words at last.

9.

Joe had decided Princeton was unwise for him after all, and Columbia welcomed him back. We began the process of gutting and remodeling a dilapidated apartment half a block away from where we'd first lived on Riverside Drive. For some months, Joe lived in and worked out of a hotel room near Columbia, while I stayed on in the Princeton house. I was deeply grateful to have in Princeton my oldest and dearest friend, Judith Gorog. I spent many happy dinners surrounded by her children, and then, once they were tucked in, further into the night with Judith and her Hungarian husband, István. They both loved good talk. They eased what would otherwise have been months of deep loneliness.

When the New York apartment was finished in late 1988, a grand wall beckoned for a Cohen painting, which we bought. Cohen stopped on his way to Europe to uncrate and stretch it, plus another for my study. The colors were astonishing, even for Harold Cohen.

Two Men on Edge stretched across the wall and dominated the room. One of my neighbors, herself a painter, came up for tea soon after we moved in. A likable woman, she lived quietly and poured all her considerable passions into her paintings. Before this massive picture, she murmured that she felt disquieted by it. Had she formed an opinion ahead of time? After all, it was "by a machine." She finally offered that it was "not quite *felt*," one of those weasely phrases that say nothing. *Too intellectual? Too perfect? Nothing else to say, so I'll fall back on "not quite felt"?* I remembered the woman I'd heard at the Art Institue of Chicago, telling us how she *felt* about physics.

Over the years, others would gaze at the painting admiringly, until we told them a computer made it. You could watch them reconsidering on the spot. It wasn't quite done by computer. It was Harold Cohen, the meta-artist, who had done it indirectly. Aaron the program was responsible for the actual image. At that point, Aaron couldn't do color, and so the image had been colored in oils by Cohen's gifted hand.

10.

Writing *Aaron's Code* was difficult; selling it was harder. I pitched editors one by one. They loved the questions the book raised; they quailed at the expense of an art book by someone unknown to them. Although the issues seemed enormous—What is art? What is thinking? What if a machine really makes art?—they resisted. *Machines Who Think* all over again. A point made later by Arthur I. Miller in his 2014 book, *Colliding Worlds*, never entered my mind: that the art establishment in 1988 was as anti-science as the humanities.

And then the publisher of *Machines Who Think,* W. H. Freeman, made a decent offer, and my heart was lifted.

The manuscript was in press by July 1990, and I wrote in my journal:

> *The adult in me expects attack from the people who hate what Harold is doing but see me as convenient scapegoat, and don't mind including me (one can*

> *hardly imagine a review saying "A lovely book about a subject unworthy of it"); by people who are open to or even like what Harold is doing, but hate an intruder on their art/crit turf. I can't win, really, so the pleasure is in the process, and we get on with the next project.*

The book came out in September 1990, and as for the many ways the critics might bash me, I needn't have fretted. The book was barely noticed. Herb Simon sent me a thoughtful, generous, and detailed review of *Aaron's Code* that would appear in some distant future in *Computers and Philosophy*. *New Scientist* informed me it was going to review the book in April, although I never saw that review. *Art & Antiques* asked me to do something for them based on the book. Jon Carroll, who for many years wrote an amusing, perceptive column in the *San Francisco Chronicle*, wrote a kind and appreciative review for the online forum The WELL, which Stewart Brand ported over to a private conference that he knew I was more likely to read. I was deeply grateful.

11.

I withdrew from the experience of writing *Aaron's Code* depleted, sad, and above all, deeply worried about my own instincts. Had John Brockman been right? The book's release certainly felt like a dead loss on every level—personal, professional, emotional, intellectual. I wondered how long it would take for me to feel whole again. Yes, I'd learned about art, but what I'd learned I'd mostly taught myself.

When I saw Cohen at a book signing in the late fall of 1991 at the University of California, San Diego, I pursued something new with him: Was Aaron a complex adaptive system? This was a term—a whole set of terms and concepts—I'd learned in September and October that year during a deeply nourishing stay at the Santa Fe

Institute, an independent think tank devoted to the sciences of complexity. The Institute was intimate enough so that you'd puzzle over a concept and walk out to find an open door where someone—above all, Stuart Kauffman (the theoretical biologist) but also Chris Langton (the originator of artificial life, A-Life), Brian Arthur (the economist), and many others, certainly including the physicist Murray Gell-Mann—would drop everything and patiently explain your puzzle to you, keep talking over lunch if you still didn't get it or if you just wanted to keep talking.

Much of my problem with the Aaron program, I'd begun to see, was the struggle to create a vocabulary for what Aaron was and did. But at the Institute, those terms and concepts already existed: they were precise, descriptive, and in daily use in the sciences of complexity and nonlinear systems. A *complex adaptive system*—the phrase I wish I'd known—was a system that began with simple rules, whose multiple layers emerged into more complex behavior, yet had no central control or leader. Such systems communicated internally both between layers and between elements of layers. Such systems changed their behavior—adapted to improve their chances of success—through learning or evolutionary processes. Aaron, blithely making its drawings, could claim countless kissin' cousins all over—in economics, physics, biology (the human brain, for one), meteorology, and many different fields.

I had lunch with Murray Gell-Mann, the Nobel Laureate in physics who knew complex adaptive systems down to his toes. On this sabbatical year of Joe's, we'd gone from a few months at the Santa Fe Institute to three months at CalTech, where Gell-Mann was on the faculty. He listened to me and nodded. Yes, Aaron was exactly a complex adaptive system, at least as it executed each drawing. Its status at the system level was dicier, but Gell-Mann cautioned me, "it's very much a matter of degree in these things."

I exhaled. I continued preparing a talk on Aaron as being "in the spirit of" complex adaptive systems. Gell-Mann had told Joe that complex adaptive systems were far more important than the quark, a subatomic particle he'd hypothesized, whose existence was only confirmed much later and for which he'd won the Nobel Prize.

Joe and I left Pasadena after New Year's 1992 and moved on to Munich, Germany, where Joe was now a recipient of a Distinguished Senior Scientist Award from the Alexander von Humboldt Foundation. That sabbatical year, first in Santa Fe, then in Pasadena, at last in Munich, restored me to myself.

12.

In the late 1990s, Cohen cracked the color problem—Aaron now chose its own colors, and they were dazzling. Aaron put colors side by side that the human meta-artist wouldn't have dared, yet the results are deeply satisfying. Cohen wrestled instead with issues of intentionality, responding to the demand we humans make of art that it not only exhibit a human touch, but that its meaning can be found in its intentionality. Thus Harold Cohen went to work on a new painting Aaron had made, perhaps changing some of the shapes, more often changing some of the colors and textures. He wrote: "It has not merely re-opened my dialog with the program, it has redefined the relationship upon which that dialog has been based." He elaborated on that in 2011: "The whole of my history in relation to computing really has had to do with a change from the notion of the computer as an imitation human being to the recognition of the computer as an independent entity that has its own capacities which are fundamentally different from the ones we have" (Estorick, 2017).

Maybe we'll all find ourselves there one day, when the world is full of intelligent artifacts. We'll begin our dialogue. And listen carefully to hear the artifacts reveal their intentions and ours. Once more, Harold Cohen was an early arrival at a place where the rest of us will eventually follow.

When I walked into a spacious and serene laboratory[3] at the MIT Media Lab in Fall 2013, nearly twenty-five years after the publication of *Aaron's Code*, I saw a stretched canvas leaning against a table, its face hidden. Because art adorns the halls of MIT (and outdoor spaces between them), I assumed the canvas was something waiting to be hung. But after a while, Kim Smith came back from lunch and got to work on another canvas on the wall. A trained artist, she was working in collaboration with Sep Kamvar, himself trained as an artist, but also an MIT computer scientist, who'd coded the art-making program. Artifacts here are a collaboration between program and humans—an artist, in Smith's case, or museum visitors, in the case of Kamvar's exhibit at Skissernas Museum in Lund, Sweden, a few years earlier. The program's instructions are both constraining and flexible, so that the finished piece has a clear structure, yet at the same time expresses the individual aesthetic preferences of the participants who contribute. "Since each step depends on previous steps," says the museum's exhibit catalog, "the result is a dynamic, collaborative piece, authored collectively by the artist [the program] and the museum visitors" (Kamvar, 2012-2013).

[3] The miniaturization of computer components has dramatically changed the ambience of computer laboratories over the last fifty years. These days they can honestly be described as serene—although intellectual excitement is anything but.

After a quarter of a century, people were ready to consider the computer as at least a partner in artmaking, if not an artist in its own right. Start-ups sell screen art. Some artists predict that screens will be the dominant medium "like canvas was for centuries," says Yugo Nakamura, a founder of one of those start-ups (Wortham, 2014). Aaron's work exhibits first on the screen, so would need no adaptation to this new world, this new kind of viewer, accustomed to screens instead of canvas.

Robbie Barrat, a Stanford researcher, took the machine-learning approach to generating paintings by AI. He fed a few thousand examples of images of landscapes into his machine-learning software until it learned how to create landscape paintings (Muskus, 2018). You might think Barrat's approach is a kind of high-level copying. However, because the software is exposed to thousands of images, it's really synthesizing, not copying. Similarly, human artists assiduously expose themselves to thousands of pictures as they're learning to make art. (Once in the 1980s I wrote an unpublished essay "Why do artists go to art museums but scientists don't go to science museums?")

Harold Cohen died quietly at work in his studio on April 27, 2016, aged 87. By then, he'd lived to see digital arts programs spring up at most major universities and art schools. Google had even established an Artists and Machine Intelligence Program, which led to an AI-based (deep learning) artwork by artist Refik Anadol to inaugurate the centennial season of the Los Angeles Philharmonic. Described as "a collage" of artifacts from the Philharmonic's history, the data on which the AI artwork is based is "millions of photographs, printed programs and audio and video recordings, each one digitized, microcrunched and algorithmically activated to play in abstract form across the building's dynamic metal surface" (Rose, 2018).

The big questions that Harold Cohen's Aaron first raised in the 1970s linger, not yet fully answered, if they ever can be.

The experience with Cohen's work changed me. In the mid-1990s, I played around with something I called "swarm stories," self-organizing stories, stories that told themselves, never twice in the same way. I tried hypertext stories, but the software was so buggy it crashed my computer again and again. Technicians took six months to discover the cause of the problem. While other writers such as Michael Joyce stuck with it, I stopped, too frustrated. But the ideas behind this software foreshadowed video games as we know them now.

The Story as the Marker of Human Intelligence?

1.

What happened next? How did the story turn out? We really want to know.

The story, Henry James once declared, is art's spoiled child (1909). By that he surely meant how readily humans surrender to stories, whether listening to, reading, or telling them. We communicate by stories—"Did I tell you . . . ?"— and make up internal narratives as self-explanations. We create stories by shaping unrelated incidents into a sequence of cause-and-effect that can be utterly false. (Danny Hillis, a distinguished computer scientist and author, has said that because cause-and-effect is just an artifact of our brain's penchant for storytelling, we should abandon the idea of it outright.) Stories are a special kind of compressed code. In a few lines, we can grasp a character's lifetime and, in a few words, be inspired, uplifted, or cast down.

The Israeli historian Yuvah Noah Harari (2015) goes further. Humans are the only species that trade *fictive* stories, he declares, which has enormous consequences. It allows us to cooperate in numbers well beyond the average of 150 individuals we can learn to know about personally, biologically, as it were. “Large numbers of strangers can cooperate successfully by believing in common myths,” Harari says, and offers religion, or nationalism, as examples. “There are no gods in the universe, no nations, no money, no human rights, no laws, and no justice outside the common imagination of human beings”. But “Telling effective stories is not easy. The difficulty lies not in telling the story, but in convincing everyone else to believe it”.

That imagined reality, that shared story, exerts great force in the world, Harari continues. Moreover, imagined realities, collective myths, and shared stories can change rapidly, adapting to new circumstances. Before the French revolution, people believed in the divine right of kings but “almost overnight” Harari says, they adopted a belief in the sovereignty of the people. Humans are open to a fast lane of cultural evolution, outstripping any other species in an ability to cooperate. By revising our shared stories to adapt to changing circumstances, humans can change their beliefs and behavior in a matter of decades, rather than waiting for the slow changes evolution brings about.[1]

2.

The foundation of stories is language. Text, which stands for words, which stand for—well, whatever they stand for—is one of our most powerful codes, and stories are one of its most powerful forms, because as humans, storytelling is one of our distinguishing characteristics.[2]

[1] Max Tegmark would phrase it differently. In his book Life 3.0 (2017), he noted that unlike other animals, humans are able to rewrite their own software.

[2] Early AI researchers recognized this. In the 1970s, Roger Schank, then at Yale, worked on programs that generated stories, which he allowed me to show to my own

I went to talk to MIT's Professor Patrick Winston, because of my own interest in stories and higher-level symbolic intelligence. Winston was a pink-skinned, affable, and trim man (not always—his website tells his tale of forcing himself to lose 60 pounds in 100 days). Winston had been at MIT since he was a freshman, loved his institution passionately, devoted much time to institute affairs, and loved to teach. He wrote a classic and best-selling textbook called *Artificial Intelligence* and, in 1972, succeeded his former dissertation advisor, Marvin Minsky, as the director of what was then known as the MIT Artificial Intelligence Laboratory, later the Computer Science and Artificial Intelligence Laboratory (CSAIL). Winston later stepped down as head of the lab but continued to teach and supervise research until his death in July 2019.

Winston's research goal, a comprehensive computational account of human intelligence, was driven by two questions. First, what computational competences are uniquely human? Second, how do uniquely human competences support and benefit from the computational competences we share with other animals?

With his colleague Dylan Holmes, Winston writes:

> Our answer to the uniquely human question is that we became the symbolic species and that becoming symbolic made it possible to become the story-understanding species. Our answer to the support-and-benefit question is that our symbolic competence, and the story-understanding competence that it enables, could not have evolved without myriad elements already in place. (Holmes & Winston, 2018)

This position is unusual—most AI today focuses on statistical

undergraduate writing classes. My students judged them cartoonish, simplistic, and I tactfully didn't say how close the computer's efforts were to theirs.

mechanisms associated with machine learning, mechanisms that shed little light on aspects of intelligence that are uniquely human, as I've pointed out. Holmes and Winston elaborate on this point:

> We believe that tomorrow's AI will focus on an understanding of our uniquely human intelligence emerging from discoveries on par with the discoveries of Copernicus about our universe, Darwin about our evolution, and Watson and Crick about our biology. These cognitive mechanisms will take to another level applications aimed at reasoning, planning, control, and cooperation. Tomorrow's AI applications will astonish the world because they will think and explain themselves, just as we humans think and explain.

Relying on work in linguistics and comparative anatomy by Robert Berwick and Noam Chomsky (2016), Winston and Holmes begin by emphasizing the *merge* operation, what they call "the *sine qua non* of being symbolic. It's the capability to combine two expressions to make a larger expression without disturbing the two merged expressions. For example, English speakers understand a bird is an animal with feathers that flies, and also understand the exception that an ostrich is an animal with feathers—a bird—but doesn't fly. Moreover, they understand from poet Emily Dickinson that "hope is the thing with feathers," which allows imaginations to think of hope as birdlike, one that probably flies (in some sense), without disturbing any of the other ideas about birds they hold. Merge gives us, and only us, an inner language with which we build complex, highly nested symbolic descriptions of classes, properties, relations, actions, and events. "When we write that we are symbolic, we mean that we have a merge-enabled inner language"

Together with the competences humans share with other species, the merge operation enables storytelling, story understanding, story composition, and all that enables much, perhaps all, of education. The merge operation also enables religion, nationalism, currency systems, human rights, and the rest of Yuval Noah Harari's list of fictive stories we tell each other (2015).

Our stories—the creating and the assimilating of them—are what make us different from other primates. They're a marker of higher-level, *symbolic* intelligence though this keystone competence could not have evolved without other elements already in place, elements we share with other species. "We developed the means to externalize our inner stories into outer communication languages, and to internalize stories presented to us in those outer communication languages" (Holmes and Winston, 2018). Thus the strong story hypothesis, first proposed by Winston in 2011: "The mechanisms that enable humans to tell, understand, and recombine stories separate our intelligence from that of other primates."

Although other animals might have internal representations of some aspects of the world, they seem to lack these complex, highly nested symbolic descriptions. Work with Nim Chimpsky, a chimpanzee who learned American Sign Language, showed that while the chimp could understand names of things and memorize sign sequences, Nim did not exhibit any merge-enabled inner language of complex, highly nested symbolic descriptions.[3] A comparison between children

[3] A few years ago, as Winston and I were meeting, he said, "Three days ago, I heard something really important. A colleague here at MIT has been able to put a probe into a rat hypothalamus. As the rat runs along a raised track, its brain waves show a sequence that corresponds to the curves in the track. As it reaches the end of the track (and its goal of food), its brain waves show that it's negotiating the track again in its brain, even though it's now standing still, eating. Moreover, sometimes it will stop on the track and play in its brain the patterns that correspond both to where it's been and where it anticipates going. It sometimes even dreams about running the track.""So this ability to imagine a sequence of events goes pretty far down the mammalian chain," I said.Winston nodded. "And we know rats are very smart." Winston further proposed

and chimpanzees shows that young humans generate novel combinations of words very freely, but Nim Chimpsky never provided evidence via signing that suggested he had this merge-enabled inner compositional capability. “Somehow we developed the means to externalize our internal stories into outer communication languages and to internalize stories presented to us in those outer communication languages. Being social animals, we started telling each other stories” (Holmes & Winston, 2018).

How did this capacity arise in humans? As Winston would tell it, it’s—well, a story. Until about 80,000 to 100,000 years ago, humans and other hominins (the group consisting of modern humans, extinct human species and all our immediate ancestors) were about the same. Ian Tattersall, a paleoanthropologist at the American Museum of Natural History, believes that sometime in that span, humans became symbolic, and parted from our other hominin cousins. Tattersall conjectures that rapid climate changes during that era forced hominins to adapt or die, and one of the most successful adaptations was the ability in a small, isolated band to manipulate symbols, in speech, in pictures, perhaps otherwise. “As far as anyone can tell, we are the only organisms that mentally deconstruct our surroundings and our internal experiences into a vocabulary of abstract symbols that we juggle in our minds to produce new versions of reality: we can envision what might be, as well as describe what is,” Tattersall writes (2014).

that intelligence is within, not behind, our input/output channels, a view generally held at MIT for at least twenty years. This means intelligence lies not in some central part of the brain, but in the use and reuse of coded vision, language, and motor systems together. A major point of agreement at an AI Summit in February 2014, convened to discuss future directions of AI, and attended by researchers from the U.S., Europe, and Asia, was that it was time for integrated systems—vision, language, and motor systems to be combined into single entities. One dissenter since has been Stuart Russell, who thinks that might make machines too smart for our own good.

Winston said: “Tattersall is a bit vague about what he means by ‘symbolic.’ He’s a paleoanthropologist. But I’m a computer scientist, and I know *exactly* what symbolic means.” (Recall early Allen Newell and Herbert Simon: symbols are functional entities. They have access to meaning—designations, denotations, information a symbol might have about a concept, such as a pen, brotherhood, or quality. The physical symbol system, whether brain or computer, can act appropriately with those symbols. (McCorduck, 1979).) Winston went on: “Then I heard Noam Chomsky talk about how we humans developed the ability to combine concepts, thus making new concepts, without destroying the original concepts.” The Genesis story-understanding program was born.

The Genesis model is being built by studying and employing the kinds of computations required to translate stories of up to 100 sentences, expressed in simple English, into inner stories. Winston and his colleagues then studied how to use the inner stories to answer questions, describe conceptual content, summarize, compare and contrast, react with cultural biases, instruct, reason hypothetically, solve problems, and find useful precedents. Nothing would go into Genesis unless it was needed and seemed biologically plausible.

Winston and his colleagues have been devoted to doing this scientifically. They’ve avoided models that are so general they can explain anything (and so are not falsifiable). Instead, their models are narrow in scope because this is only the beginning: Genesis, its builders say, is analogous to the Wright Brothers airplane of 1903.

Genesis has learned from summaries of plays, such as Shakespeare’s *Macbeth*; fairy tales, such as *Hansel and Gretel*; and contemporary conflicts, such as the 2007 Estonia-Russia cyberwar. As Genesis reads simple concise stories, it connects causes to effects and means to actions, sorts membership in classes, and uses inference to elaborate on what is written. It reflects on its reading, looking for concepts and concept patterns that allow it to make abstractions. Thus Macbeth harms Macduff, and Macduff wants revenge, a word that doesn't appear in the summary Genesis has read. The system can do more: it models personality traits and anticipates trouble. It aligns similar stories for analogical reasoning (using an algorithm from molecular biology!). For example, Genesis finds clear parallels between the onset of the Arab-Israeli War and the Tet Offensive in the Vietnam War. "In both cases, intelligence noted mobilization, intelligence determined that the attackers would lose, intelligence determined that the attackers knew they would lose, intelligence concluded there would be no attack, whereupon the attackers promptly attacked. Retrospectively, there were political rather than military motives" (Holmes & Winston, 2018).

"I'd go beyond that, though," Winston told me, "and say the most important concepts to combine are event descriptions. We combine event descriptions into larger sequences; then we move backward and forward in remembered sequences. With that ability, we can tell stories, understand stories, and combine old stories to make new ones. That, I think, constitutes part of the answer to the question of what's different about us."

So humans developed the capacity for a complex inner story—possibly owing to a completed anatomical loop, incomplete in other animals, Berwick and Chomsky hypothesize—and then the ability to externalize those inner stories, and internalize stories presented to us, "and because we are social animals, externalization and internalization had a powerful amplifying effect." In other words, what Yuval Noah Harari names as the unprecedented ability to cooperate, owing to shared stories.

Storytelling, Winston believed, makes it possible for humans to construct elaborate models of ourselves (possibly consciousness?) and the world outside us. "If we're to understand human thinking," he speculated to me, "we must model that story-manipulation, model-enabling capability. In the end, that's what makes us different from species that have plenty of just-do-it, *and* simulation capabilities, but whose story manipulation capability, if any, is on another, much lower level."

Being symbolic, Winston went on, allows humans to have an inner language that supports story understanding, the acquisition of common sense from perception, and the ability to communicate with others. Of course, we share much with other animals, too, which remains to be fully understood.

Although Winston appreciated efforts that have led to outstanding engineering, such as Rodney Brooks's robot insects, not to mention Brooks's wildly successful Roomba robot vacuum cleaner, Winston was personally more interested in what he calls the science side of intelligence, symbolic capacities. The founding fathers of AI also believed symbolic capacities were central to intelligence. Winston thought the way forward was to ask "better, biologically inspired questions." Good science informs good engineering or applications.

In Genesis, Winston believed he'd departed from early AI's view of what it meant to be symbolic, that being symbolic meant only logical reasoning, and nothing else mattered. "I think that reasoning is recipe-following. Recipe-following is only a special case of story understanding," he said to me.

Yes, the earliest AIs embodied logical reasoning, but given how much Newell and Simon, for example, honored and practiced storytelling themselves (remember "The Apple" and "Fairy Tales"), they never believed their reasoning-based programs were all there was to intelligence. Each of them said so explicitly. (As we saw earlier in this book, Simon called logical reasoning a "small but fairly important subset of what's going on in mind"). Early AI was based on what was immediately accessible to cognitive psychologists in the mid-1950s, those thinking-aloud protocols of reasoning, as subjects tried to solve problems, coupled with the primitive computing technology of the time. That you might one day be able to read human or even rat brain waves, much less exhibit the electro-chemical behavior of the brain, was beyond anything at the time.

Newell and Simon declared explicitly that there was much more to thinking than what they could then simulate, and both would be comfortable, I think, with Winston's emphasis on storytelling as an indisputable marker of human intelligence.

Winston and his colleagues worked with neuroscientists and psychologists to push these ideas further. Winston did so as an investigator participating in the Center for Brains, Minds, and Machines, an MIT-Harvard interdisciplinary group of computer scientists, neuroscientists, and psychologists that meets regularly to exchange ideas and findings about cognition.

Genesis does not aim to advance the state of the art in question-answering, as IBM's Watson does, for example. Its creators intend to devise and build a plausible and scientific account of human story understanding, showing how a story-understanding system is able not only to answer questions, but also describe conceptual content, summarize, compare and contrast, react with cultural biases, instruct, reason hypothetically, solve problems, and find useful precedents. (This reminds me of the original Logic Theorist, which wasn't built to be a killer logician, but to model how humans proved theorems in logic.)

The simple substrate of Genesis supports many competences, Winston declared. Some examples: Genesis answers questions about why and when, models personality traits, notes concept onsets, anticipates trouble, and can re-interpret stories with controllable allegiances and cultural biases. (It first views the cyberwar between Estonia and Russia as the aggression of a bully, from the Estonian point of view, and then as teaching a lesson from the Russian point of view.) Another example includes Genesis' ability to persuade.

Winston viewed story understanding as foundational to human intelligence. To understand and model it in detail is a significant step toward constructing artificial intelligence. For now, Genesis reads and demonstrates all these capabilities only around stories that are adapted for it. The model cannot understand stories written by people for people. Critics complain that Genesis should learn, not be instructed, although most humans must be instructed—by their parents, by their schools, by experience—in many of the issues Genesis confronts.

My life has been shaped by stories. The most intimate and enduring transaction of my life has been to transfer an outer story to an inner one, an inner to an outer one. My mother read Enid Blyton to me, and so momentarily I became one of Blyton's plucky children. But when my brother and sister, twins, were born, I was on my own. By then, I could read and took up my mother's copy of unexpurgated

Grimm's Fairy Tales, transmuting horrors of cruelty and even death into inner stories, which would to teach me far more about real life than the denatured "children's books" I encountered when we arrived in the United States. Much later I became for a while Dorothea Brooke, Isabel Archer. I began to transform my own inner stories to outer ones, as I have in this book.

So I observe the steps toward story understanding that Winston and Holmes propose as precise and explicit: knowledge acquisition, concept formation, analogies to other stories Genesis knows, the ability to reason and summarize, to persuade a reader from a given cultural point of view, and more. Much work is to be done, but Genesis is only Kitty Hawk. Genesis is only a first draft.

3.

Understanding stories takes many forms. Oren Etzioni, who in 2013 became the first head of the new Allen Institute for Artificial Intelligence in Seattle known as AI2 (founded by Paul Allen and mostly, but not entirely, funded by him), has also been long at work on text understanding. "Why text?" I asked, knowing that so many of his colleagues are working on other kinds of perception—machine learning, for example—as a route to intelligent machines. "For the same reason Willie Sutton went to the banks," Etzioni laughed. "The banks are where the money is. Text is where the knowledge is—all over the world."

AI2's approach is called open information extraction. It isn't just fact-finding, but fact understanding—finding both knowledge *and* meaning in text.

AI2's efforts include a series of programs that can pass fourth-grade, eighth-grade, and twelfth-grade tests in science, language arts, and social studies. The programs must meet explicit benchmarks. For example, in fourth-grade arithmetic, to tease out the essence of word problems requires not only the ability to think through the problems, but also real world knowledge—lifespans, what animals are, and so on. When we succeed at that, Etzioni says, we'll only have an artifact. Which, of course, can be built upon.[4]

The second goal at AI2 is for the common good: a better scientific search engine, called Semantic Scholar, that can "understand" and search semantically instead of using keywords in context, like Google Scholar.[5] How will success be measured here? By how users behave: what they ask, how often the system is used, and whether and how often users return.

In 2017, AI2 announced a new project: giving computers common sense. Project Mosaic (first called Project Alexandria as a tribute to the great ancient library) builds on earlier programs the Institute has been working on, including machine reading and reasoning (Aristo), natural language understanding (Euclid), and computer vision (Prior), to create a new unified and extensive common-sense knowledge source. Mosaic will also draw on crowd-sourcing.

AI2 researchers work closely with the Allen Institute for the Brain because the long-term goal is to discover and define what intelligence is. "This is the grand question. It will take a long time," Etizioni says. Meanwhile, AI2 will not only work on these specific shorter-term goals, but also sponsor distinguished investigator awards, stipends to

[4] Current papers of AI2 are posted on the Institute's web site (http://allenai.org) so you can judge the progress of research.

[5] Access Semantic Scholar at https://allenai.org/semantic-scholar/

individuals who are eager to go beyond incremental approaches to AI and think in larger, more comprehensive terms.

This voracious consumption of text in order to know and understand is underway (with variations in methods and ultimate goals) at all the major computer firms—IBM, Google, Microsoft, Apple, Facebook—and many research sites. Each effort takes a different approach. Carnegie Mellon's Nell (Never Ending Language Learner) program is a machine-learning project that "reads," or extracts facts from, text found in hundreds of millions of websites, to which it assigns different levels of confidence. It attempts to improve its competence so that it can learn better tomorrow, extract more facts more accurately. You can visit Nell's website (http://rtw.ml.cmu.edu/rtw/) and see the categories it has extracted facts from, and whether you agree. Another program at Northwestern can convert numerical data (such as sports scores or profit-and-loss statements) into stories: a sports story about your youngster's Little League game or a story to help a franchise manager understand why the branch across town is doing better.

4.

In connection with machine learning, I've mentioned the largely unspoken assumption, maybe hope, that as machines accumulate the abilities that correspond to lower human faculties, the higher faculties will inevitably, or magically, emerge. We call this level of higher faculties *symbolic intelligence*, and emergence seems to be how it happened with humans, so why not?

Lots of reasons why not, and they aren't necessarily about better hardware—although they could be about better software. Google's Giant Brain seems to have as many connections as the human brain, but requires megawatts of power, whereas we're still smarter in some ways with only 20 watts. (Although what "smarter" means is problematical.) The abstract kinds of language and thinking that might have emerged some 80,000 to100,000 years ago from a single hominin group was a winning elaboration of the relatively simple communications our hominin cousins already had. As I've noted, Ian Tattersall (2014) conjectures one small, isolated group produced and sustained symbolic capabilities, and because they were few and isolated, the genes were allowed to flourish.

Leslie Valiant has said it's impossible for human coders to do what machines eventually must do automatically to achieve intelligence (probably so), and Eric Baum posits an underlying structure in the world that is detectible and amenable to a compressed representation. Perhaps such a structure exists. If so, for millennia it has been science's grand quest to find it. We'll see.

In 2015, Kathleen M. Carley, a professor in the School of Computer Science at Carnegie Mellon, presented "Will Social Computers Dream?" at a symposium, where she shared her strong belief that, as interesting and capable as machine learning is, it can't achieve human-like intelligence without social cognition, the ability to reason in a socio-cognitive-emotional fashion. These are a set of procedures and behaviors humans follow to reason about and respond to the world from both "social collective" and "individual affective" viewpoints. Social cognition is partly physiological, partly learned, requiring the actor to be in a rich socio-cultural environment, engaging in real interactions with multilevel actors and with multilevel and competing goals, histories, and culture. To give complete social cognition to computers is complicated, she says, and is unlikely to happen in the next fifty years. But computers with partial social cognition are likely to emerge before then, with generally positive advantages for humans themselves (Carley, 2015). Perhaps they will also have the kind of distributed intelligence Winston believed is needed for genuinely human-like intelligence, not in some central part of the brain, but using and reusing coded vision, language, and motor systems, which will lead to inner and outer narratives.

If video games are the new storytelling, video game designers also have high ambitions to incorporate social cognition into games (Stuart, 2018). First, they want to remove the interface—that is, get rid of knobs and joysticks and allow participants to play the game with voice commands. Developments in animation and motion capture will soon allow more than words to present the nuances of a character's behavior. Rigid narrative conventions will be replaced by AI-driven *reactive systems,* intelligent games in which the game engine develops a sense of dramatic control that enables it to decide the best moment for a player to meet another character. Because the stories will be open-ended with new possibilities added daily, video-game designers will have to learn how to tell stories that evolve over months or years. Perhaps most ambitiously, the best narrative designers want to develop stories that speak to cultures all over the globe. Margaret Stohl, a successful designer, mentions how loneliness afflicts so many humans, and adds: "People don't think of video games as emotionally progressive, but as online communities thrive around them, that's a chance to be part of something".

It remains to be seen whether general intelligence can be achieved or whether the white-hot research of machine learning will lead to the spontaneous emergence of symbolic intelligence (thinking slow). But sooner or later, I believe we'll be facing human-level general intelligence in our machines, except that their powers will be faster, wider, and deeper than ours.

Of course as with all aspects of AI, a troll lurks under the bridge. OpenAI, a nonprofit research group in Silicon Valley, has created GPT2, a text generator that is so good at writing news stories and fiction that the organization has decided not to release it yet (though that's the point of OpenAI) because the potential for malicious use is so great. Fed just a few starter lines or paragraphs, the system takes up the narrative and continues with a story so plausible (complete with fictitious quotes, if it's a news story, from major figures concerned with the story's topic) that for a reader to detect whether the story is real or fake is nearly impossible. GPT2 has trained on very large data sets, and can be tweaked further to be positive or negative. OpenAI researchers are testing the system, to find out what it can and cannot do, especially maliciously (Hern, 2019). Meanwhile, the story—whether poem, novel, five-season TV series, or advanced video game—appears to us in combinations of words, images, and music, each a kind of code, each a technology of compression. (So is mathematics. So is music notation. So is computer programming.) A new field is beginning to view words and text (and music, and images) in just such terms. In a gratifying closure of my life's circle, this new field is a wedding of the Two Cultures called the digital humanities.

The Digital Humanities

A Great Moment in Cultural-Historical Transformation

1.

In the fall of 1960, about the same time I heard C. P. Snow deliver his *Two Cultures* lecture at Berkeley and was introduced to AI, I was taking a course in Italian Renaissance literature. When I visted my professor's office for a consultation, he told me he'd nearly finished a concordance to Petrarch. After years devoted to Petrarch's every written word, with 3 x 5 cards crammed into shoeboxes on shelves, the desktop, the floor, he was jubilant.

"Too bad you didn't use a computer," I sniffed, surely spoiling his pleasure a bit. But I was the newly converted and a pain.[1]

Over the years, I grew more tactful, but as this history reveals, my efforts to bring AI, and computing generally, to the attention of humanists—colleagues when I was teaching at a university, New York editors as I became a full-time writer, librarians, other writers when I was active in the PEN American Center, the authors' freedom

[1] In 1949, the Jesuit scholar Roberto Brusa worked in collaboration with IBM to create an automated approach to his Index Thomisticus, a computer-generated concordance to the writings of St. Thomas Aquinas, but he needed access to mainframes, so flocks of scholars weren't able to follow his example.

of expression organization—were mostly futile. “This could be important,” I’d say. If they listened, which was seldom, they scoffed. At best they wanted to redline me into a narrow cell called science writing. That computers, never mind AI, might have a larger significance seemed absurd to them.

Computing surely seemed absurd because so many of them were at such a willed distance from it. As a new board member, I’d walked into the PEN American Center offices in the early 1980s to find typewriters, and membership records in, yes, shoeboxes. Before Vartan Gregorian knew my name, he took me aside at a New York City Library fundraiser and asked for a donation “before computers take over the library.”

I couldn’t expect humanists to learn what had taken me more than a decade’s hard work to understand. But I faulted them for not troubling to ask whether this could be important.

So I drifted from the formal humanities. That didn’t mean I stopped reading literature or history, listening to music, going to galleries, or pondering philosophical questions. Instead, like millions of others, I did all those for the deep human joy of it. Novels and poetry, history and biography, music, the visual arts, the best way to live a good life: all these represent human concerns at one immediate, endlessly compelling level of existence.

Biologist Edward O. Wilson (2014) reminds us that our fascination (maybe obsession) with each other is wired in, an adaptive characteristic of our species that has helped allow us to prevail. It’s the evolutionary excuse for our intense preoccupation with ourselves, eternally evaluating one another in “shades of trust, love, hatred, suspicion, admiration, envy, and sociability” These are the traditional tasks of the humanist, although humanities scholars go further, by examining in depth how our self-fascination manifests itself in works of art, be that painting, literature, music, religious beliefs, or history.

Questions around the humanities are embedded within other questions. How did we get here? What makes us unique in the biosphere? In the cosmos? (If we are?) How is collective human behavior different from individual behavior? What accounts for the many contradictions in our behavior? Why are we noble? Why self-destructive? At a different level from art, these questions are so far more precisely answered by science. The best answers take combined approaches.

From the beginning, mystics have understood the significance of such questions and offered us religious myths. But as we saw in Chapter 23, Jaron Lanier argues we've confused the science and technology of AI with mythology, reinventing AI not as something that requires thoughtful deployment, but as a divinity to be feared. He's right, and it's regrettable. Mythical answers of any kind no longer wholly satisfy a secular and scientific age. Instead, human and machine intelligence are now seen in a grand computational framework called *computational rationality*: a converging paradigm for intelligence in brains, machines, and minds.

Poised at the beginning of adult life, that young woman who drank in the Two Cultures lecture so thirstily was to spend her—my—adult life longing to reconcile the two cultures that I loved so much. My intuition pushed me slowly toward speculation, then conviction, that this symbol-manipulating device called the computer, especially this branch of computer science called artificial intelligence, would illuminate human intelligence in important ways. Yes, it was engineering, it was science, yet it might reveal some secrets that had so far evaded us. It might do more. I could only guess. Maybe hope.

What I couldn't know—and it's significant—was that the lecturer who set Anglo-American letters disputaciously ablaze with his Two Cultures thesis in the middle of the 20th century had the same yearning. Deirdre David (2017), in her biography of novelist Pamela Hansford Johnson, Snow's novelist wife, notes how Snow and Johnson's love and marriage was very much grounded in writing stories and how profound Snow's yearning had always been to be taken seriously not as a scientist but as a writer. Doing science was nearly accidental for him, but came to provide the milieu he could write fiction about. He too wanted to conjoin the Two Cultures. Perhaps his talk stuck with me for so many years because, without knowing it, I responded to exactly that yearning.

Over the decades, computers have penetrated our lives. AI moves slowly toward human-like behavior and in some cases betters it. Natural language processing improves, likewise automatic translation, and those improvements offer important techniques to linguistics and new ways to describe phonology, understand human language processing, and model linguistic semantics (Hirschberg & Manning, 2015).

AIs have learned how to interpret and respond to human facial expressions (hence human emotions, which some programs can even anticipate) and are learning how to interact safely and inoffensively in human spaces.[2] Composers have brought the digital to music in unexpected ways; movies have been voracious, pushing digital visualizations ever further; art that owed its being to the computer

[2] See especially the work of Julie Shah at MIT.

hangs on my walls—either original and grounded in AI, like Aaron's, or images transformed by a human sensibilility, like Lillian Schwartz's. Younger artists demonstrate that advanced video games will be the next great storytelling medium. Storytelling, as we've seen, is one of the great markers of human intelligence.

But the idea that the digital might also be touching literary scholarship or other parts of the humanities was news that came to me fitfully. I'd hear a talk. I'd read something in passing.

Now, as knowledge is being moved at breathtaking speed from human texts, images, and skulls into software, attention is at last being paid. Formal programs in the digital humanities have been established in nearly all the major American and European universities, in the conviction that this is a great moment of cultural-historical transformation. Because the entry fee is relatively low, schools of more modest means can also participate.

Borders around the digital humanities are porous, nearly nonexistent, because settlements can be established anywhere that humans and computers can go. Quantitative analysis is growing more important in the humanities, and text as the primary repository of human culture is challenged. Text is blended with, even transformed into, the graphical, the musical, the numerical. But the purposes are sweetly familiar: to know and understand as deeply as possible what humans feel, know, and do. Recall Leslie Valiant's claim: "Contrary to common perception, computer science has always been more about humans than about machines." (2014).

2.

In early 2015, it seemed fitting to return to the Berkeley campus where, more than half a century earlier, the Two Cultures, along with artificial intelligence, had ambushed me simultaneously. Compared with some institutions, Berkeley wasn't far advanced in the digital humanities then, but a small, lively program existed, supported by a $2 million grant from the Mellon Foundation. An enthusiastic team, which reports to the Dean of Arts and Humanities, has slowly seeded the campus with modest projects. Claudia von Vacano, who heads this program, told me their strategy: as one humanities scholar transforms his or her work, colleagues will watch, and become inspired.

The Dean of Arts and Humanities, Anthony Cascardi, himself a professor of comparative literature, has crisp reasons for the humanities to be involved in the digital world. First, the computer can organize, sort, investigate, and navigate great swaths of information and bring access to content that is now inaccessible, he told me. (It's a longtime scientific and computing axiom that *more* is *different*, that quantity changes quality.) Then, the computer allows intersections with different media, creating new research that in turn creates new content. In short, this is where the world is, and the humanities need to be there, too.

The successes and challenges of Berkeley scholar Niek Veldhuis, professor of Assyriology, typify some of the early struggles of the digital humanities. Sumerian is a linguistic isolate: it's seemingly related to no other known language. However, Akkadian, a Semitic language, was spoken contemporaneously with Sumerian, and from context, scholars are painstakingly deciphering what written Sumerian—the marks on cuneiform tablets—might mean by comparing them to known Akkadian words.

Some decades ago, a Sumerian dictionary was undertaken, but after 25 years, only four codex volumes had been produced, going from A to the beginning of B. Even by scholarly standards this was frustratingly slow, so instead, Veldhuis is trying an intermediate approach. He scans cuneiform tablets and pulls out words to post online with glossaries for other scholars to examine and interpret. This is not yet a dictionary, but its raw materials.

The project was actually begun by one of Veldhuis's former professors at the University of Pennsylvania, who made images of tablets on a simple flatbed scanner. Veldhuis laughed as he recalled to me one early problem. A technologist said: how cool would it be if you could look at these tablets in high-definition 3D, turn them, see each side. So that was begun, but the process is so expensive that far fewer images have been produced, and none is available for free online. The tradeoff between gorgeous, up-to-the-minute technology and cheap sufficient technology is apparent. In the future, it might be possible to download many gigabytes cheaply, Veldhuis mused, but not now.

He also devised a projector that allows his students to explore the image of a given Sumerian word in the seminar room together. The projector is a pedagogical tool to help refine their research techniques quickly, justify reasoning before the group, and speculate about other possibilities. "We all watch the process of research, asking how do you verify? What tools do you use? New questions arise from new technologies," he added. "That's heartening. We need to grab opportunities."

Lately he and his colleagues have begun to map the communities and social patterns of Mesopotamia, a relatively new project in the field. They analyze digitized data from early clay tablets that contain inheritance documents or sales contracts to reconstruct the social relations of ancient Mesopotamia, producing a graphical representation that tells us much more about how people actually lived and interacted.

"My expectations," he mused, "were that the technology would be hard—but no, it's the cultural differences that are hard. This is the point of the 3D versus the flatbed technologies. How much technology is necessary? A colleague wants to produce an online version of *The Egyptian Book of the Dead*, with images and commentaries on those images. Technologists will tell you this is easy. The problem is, which system will work best for what she wants to accomplish? She needs many conversations with technologists to settle on a system that does everything she wants to do."[3] Never mind the problems of how systems grow obsolete and are replaced.

Veldhuis sees the digital humanities as not only developing technique, but as social engineering. More humanists need to know what can be done and learn how to do it. Among his challenges is helping his colleagues to understand that work he posts online is provisional. This online work isn't as certain as the material he'd publish in a scholarly paper but is there to be examined, queried, and tested against hypotheses. In his field, this approach is highly unusual—scholars usually publish only what they are certain of.

Before I left his office, Veldhuis indulged me in a brief conversation

[3] Rita Lucarelli, Assistant Professor of Egyptology in Berkeley's Department of Near Eastern Studies, researches religion, magic, and funerary culture in ancient Egypt. Her Book of the Dead digital project is now underway, and focuses on creating highly detailed, annotated 3D models of funerary objects to better understand the materiality of the Book of the Dead texts.

about the origins of writing, a history of experimentation, he said: some things worked, some didn't. (Generate and test, computer scientists would say.) "But finally in the fourth millennium BCE, cuneiform took hold, partly driven by urbanization and the complexity of urban life, with its specialization and larger population, this despite apparent political upheavals. We don't know for sure because no political records exist for that time. Written records at the beginning were essentially accounting records: *two goats and a sheep.* Cuneiform prevailed thanks to its flexibility, and within a century, could express political statements. *So-and-so is King.*"

He dug into a drawer and brought out the real thing, cuneiform tablets he kindly allowed me to hold in my palm—small incised clay fragments the size of a domino, the profoundly thrilling distant ancestor of the words I write.

The ebullient Elizabeth Honig, an associate professor of art history and another scholar using the digital humanties at Berkeley, specializes in the oeuvre of Jan Brueghel, son of Pieter and father of Jan the younger. Early in her career, Honig realized that with so many extant works, knowledge about them was scattered in many human heads around the world. Pooling that knowledge would be extremely useful. Encouraged by a senior European scholar, she constructed an informal wiki that other Brueghel experts could contribute to and consult. This led to an early Mellon Foundation grant to put together a proper website and digitize many works. Luckily, her life partner is a computer scientist and helped her through some of the difficulties in setting up the website. Soon she received another grant for course development and can give students modest course credit for contributing to the website. She too sees this as a pedagogical tool that helps students develop their art historian skills.

One student's task has been to identify the patterns that appear in Jan Brueghel's paintings. The painter and his studio often employed repeated patterns of travelers in a group, windmills, or other such images, sometimes using the equivalent of stencils. To identify these is one way of authenticating a painting or a drawing. Scholars also find it useful to detect how these patterns vary in otherwise authenticated works. Honig hopes someday to be able to compare and contrast works automatically, answering whether one painting riffs on different compositional elements of another painting.

Eventually the Jan Brueghel website will have all the data the original wiki contained and an underlying database that allows scholars to trace the events and reasoning that drove authentification in the first place. The site will contain a timeline and maps and become a public utility. The website has presented unusual challenges. "Dealers can be unscrupulous," Honig told me. Because at least one dealer hacked the website, put up an image of a fake Brueghel he was trying to sell, and added fake provenances, the site now requires anything added to it to be traceable by any scholar.

Art historians are taught that a very high-resolution, black-and-white image of a work is better for analysis than a color image—details like brushwork are more apparent, for instance. When Honig studied such a photo of a painting in a private English collection, its odd brushwork made her question its authenticity. This made a London dealer nervous. He was selling the painting to a private collector in Shanghai, and Google Analytics on Honig's website showed that parties on both sides were consulting the website repeatedly. The website had become central to the authentification process. (Eventually the painting was authenticated.)

But like Veldhuis, Honig too has faced other challenges this early work raises. When we met, the website had already required five years of work, and useful as it might be, no scholarly credit accrued to her in terms of professional promotion, which normally rests on peer reviews. Thus she's had to begin writing a book, interpretive of Brueghel's art, as distinct from the website, which is solely factual.

Moreover, she's concerned about the ephemeral nature of art websites. Who does them? Who will keep them up? How long will they last? On the positive side, younger people are at home online and know how to navigate, which requires different skills from reading a monograph. Despite the challenges, she knows exciting things in art history are being done. As one example, she named the Bosch Research and Conservation Project,[4] which is devoted to the work of the fantastical Dutch painter Hieronymus Bosch and designed and organized for art historians and the way they think.

Like Veldhuis, Honig sees ambitions that are sometimes too steep for current possibilities. Basic, open-source tools are needed, a point that Anthony Cascardi, the Berkeley Dean of Arts and Humanities, also made to me. Berkeley is hoping to solve that problem with a campus-wide project that provides such tools. I'm put in mind of the early history of programming languages, when everyone made up their own special-purpose language. More powerful, flexible languages eventually came to dominate, and that will happen with digital humanities tools, too. But for now, problems persist.

[4] You can view the Bosch Research and Conservation Project at http://boschproject.org

Under the eucalyptus trees outside the Center for New Music and Technology, I met with Edmund Campion, a genial and enthusiastic composer and member of the music faculty, to talk about his uses of the computer in musical composition. (February under the eucalyptus trees! I'd just left a completely icebound Manhattan.) Campion began:

> Music has lots of data, and thus has always responded to new technologies: Mozart loved a new instrument called the clarinet; Beethoven introduced the trombone into several of his symphonies. Thus I don't think of myself as a digital composer. I'm a composer. I'm doing what composers in the past have always done, taking advantage of whatever my times offer me. Any composer needs to assess a site, an orchestra, or a new instrument, or a computer: what can I generate from this? We need to mine systems for their creative potential, and find a sense of alignment with the possibilities of our instruments. I leverage every possible means at my disposal.

Campion is unusual in having been academically trained, and yet believes he got his real musical education by playing in rock bands and progressive jazz ensembles for real audiences. This makes his music and his approaches very different from the classical music of most of the second half of the 20th century, which he calls "nearly unlistenable." Composers who began with computers were in denial about this, he says ruefully. About his time at IRCAM, the French center of avant-garde music, he said:

> The issue was *ecriture* vs. sculptural electronic music. We were asking, "What is a note? It isn't at all clear." The 18th and 19th century composers got all that under control, but then Schoenberg opened it up to all sounds again.

But this preoccupation and its experiments overlooked the listener.[5] Music, Campion strongly believes, is a social contract:

> You must have listeners—you must be part of a community. I don't mean popular, as such, but there must be a community of listeners who respond to your music. I think of music as a lattice—there are all kinds, and they're all wonderful. But music cannot happen in a vacuum. It needs a community.

Moreover, now is a time of collaboration, which the computer enables. "I trade files with video artists: is this music working for your video? Yes? No? What changes will make it work?" Campion said. He stopped for a moment and then continued:

> This brings me to the digital humanities problem. I'm wary about the lack of engagement on the part of the humanities that might result in a massive forgetting. Young people want to use the new technologies, and if they don't have guidance, if they have no feed-forward from those who have gone before, that's a terrible loss. My role as a cultural agent is to make connections between the past and the future. My students are very accomplished with the tools; they've grown up with them, which I didn't.
>
> But they have no sense of stepping off from where their predecessors were. If I don't introduce them to the past, they don't know about it. That's very difficult, because the technology has changed so much over the last half century that nobody has all the resources to bring the old stuff to the new platforms. It's all but lost.

[5] John Adams's engaging memoir, Hallelujah Junction: Composing an American Life (Farrar, Straus and Giroux, 2008) covers this conflict thoughtfully. Adams, his young ears full of early rock music and jazz written for audiences and meant to stir, abandoned the sere wastelands of Boulezian dicta to write music to be listened to. He was an early adopter of electronics, and when he couldn't afford a proper synthesizer, he built one out of used spare parts.

Although Campion is talking about music, the same peril is everywhere as the digital divide grows and predigital scholars fail to engage with the young. These are the unsurprising growing pains as the great structure that I talked of earlier begins to rise, that Hagia Sophia called computational rationality, encompassing, connecting, uniting and defining intelligence wherever it's found, in brains, minds, or machines

Three years after my initial visit to these Berkeley digital scholars, Anthony Cascardi, that lively dean of the Arts and Humanities, reintroduced me to Charles Faulhaber. (We'd met earlier when he was head of the Bancroft Library at Berkeley, which specializes in American studies.)

Faulhaber is a professor emeritus of Spanish literature and in his retirement can now devote himself fully to a project he's been at for a while. It's called PhiloBiblon[6] (after a medieval text describing the perfect library) and is an online bio-bibliographical database of medieval texts of the Spanish, Portuguese, Galician, and Catalan languages. It documents the kinds of texts that will serve as data for the *Diccionario del Español Antiguo* being resumed by the Real Academia Española after a hiatus. This will be a Spanish equivalent of the *Oxford English Dictionary*, that is, the notation of a word from its first appearance in text with examples of usage over time. PhiloBiblon will also serve other lexicographical projects, critical editions, and other text-based projects focused on medieval Spain.

Any scholar may access it, but Faulhaber's ambition, he tells me, is to move the entire database from the present Windows format to the World Wide Web, so that it can take advantage of the semantic web

[6] You can find PhiloBiblon at http://bancroft.berkeley.edu/philobiblon/

capabilities. At the moment, only one person at a time can add to the PhiloBiblon database. “We can provide web access to the data, but it is not an elegant process,” Faulhaber told me. His goal is to put the database on a server where any authorized user would have access to it, from anywhere in the world, in order to add data. “An editorial committee would vet all changes to ensure that they conform to our standards—a classic crowdsourcing application.”

Thus scholars in libraries in Spain, Italy, France, England, or Russia could add data from real-time inspection of primary sources. This would eliminate the single greatest bottleneck in maintaining and expanding that data. “There is no substitute for first-hand inspection of primary sources, but the semantic web will make it much easier to look for those sources,” Faulhaber says. “Every day, libraries all over the world are adding data about their holdings as well as digitizing them. Finding these new materials on the web is a hit-or-miss proposition right now.” Currently the only way to add these data to PhiloBiblon is by manual cutting and pasting. “We’ve done it that way for forty years, but the semantic web makes it possible to automate this process.”

The semantic web is a network that automatically collects texts so that *meaning* can be teased out of them, to follow the progress of a given word through its evolution in the language. This would be impossible without AI. Picture this, and contrast it to the lone scholar’s labor to produce a concordance of Petrarch’s works, which opened this chapter.

In 2017 faculty colleagues in engineering and computer science across the Berkeley campus issued a technical report, *A Berkeley view of systems challenges for AI* (Stoica et al., 2017). Although this was intended for computer science and engineering readers, some salient points applied to the digital humanities, especially as they began to employ AI. A systems approach would be necessary, crossing disciplines, and such applications needed to support continual or life-long learning and never-ending learning. The report notes that Michael Mccloskey and Neil Cohen define continual or life-long learning as "solving multiple tasks sequentially by . . . transferring and utilizing knowledge from already learned tasks to new tasks while minimizing the effect of catastrophic forgetting." Never-ending learning, however, is "mastering a set of tasks in each iteration, where the set keeps growing and the performance on all tasks in the set keeps improving from iteration to iteration." Other challenges include adversarial learning which occurs when malefactors compromise the integrity of applications—that unscrupulous art dealer who inserted fake merchandise onto the Bruegel website, for instance.

Meeting such challenges is probably beyond an individual humanities scholar's abilities, but that scholar needs to be able to ask the hard questions and insist that cooperation among different kinds of experts is essential.

Humanities Now and Forever

1.

The humanities willingly transform themselves, embracing the computer enthusiastically. Their embrace—whether they know it explicitly or not—incorporates principles Jeannette Wing has for more than a decade called "computational thinking." Wing is the Avenessians Director of the Data Science Institute at Columbia University and a professor of computer science there. Computational thinking, she says, is a universally applicable attitude and set of skills that everyone, not just computer scientists, is eager to learn and use (Wing, 2006).[1] Computational thinking is part of the "durable intellectual content" that Mary Shaw, the Alan J. Perlis Professor of Computer Science at Carnegie Mellon, has called for, that makes computer science a science, beyond the technology of the moment (Togyer, 2014).

Computational thinking builds on both the power *and* limits of computing, whether executed by a human or by a machine. These methods and models give us

> . . .the courage to solve problems and design systems that no one of us

[1] Wing has subsequently elaborated upon her ideas in many articles.

> would be capable of tackling alone. Computational thinking confronts the riddle of machine intelligence: What can humans do better than computers? and What can computers do better than humans? (Wing, 2006)

Computational thinking is solving problems, designing systems, and understanding human behavior. It includes a range of mental tools that reflect the breadth of the field of computer science. So we ask of a particular problem: How difficult is it to solve? What's the best way to solve it? These are questions that computer scientists can offer precise answers to. We ask, Is an approximate solution to the problem good enough? "Computational thinking is reformulating a seemingly difficult problem into one we know how to solve, perhaps by reduction, embedding, transformation, or simulation." It's thinking recursively; it's parallel processing; it interprets code as data and data as code. It judges programs not just for correctness and efficiency but for aesthetics. It judges a system's design for simplicity and elegance. And more. (Wing, 2006)

Computational thinking is about conceptualizing, not programming. It demands thinking at multiple levels of abstraction. Abstraction and decomposition are used to tackle large complex tasks or to design large complex systems. Computational thinking chooses an appropriate representation for a problem or models the relevant aspects of a problem to make it tractable. "It is planning, learning, and scheduling in the presence of uncertainty. It is search, search, and more search . . ." Wing writes.

Above all, computational thinking is a fundamental. It's not a rote skill. It's about ideas, not artifacts. It's a way that humans, not computers, think. It complements and combines mathematical and engineering thinking. It's for everyone, everywhere, an intellectual adventure that will be commonplace to human thought in the future.

2.

To calculate the present extent of digital humanities projects would be hopeless—they spring up daily (and sometimes languish as quickly). A Google search will lead you into this vast territory. It includes dynamic maps of the encounter between European and indigenous peoples, multimedia projects exploring 19th century music in Victorian literature, and archives of performances of Greek and Roman drama. A Stanford project examines the Roman world, considering travel patterns and their effects on governance, art, and literature, and lays it out in graphic terms for you. The Homer Multitext Project,[2] housed at both the University of Leipzig and Tufts University, exhibits multiple texts of Homer's work from all over the world, for scholars sitting in their studies to access and compare. Visual reconstructions abound of medieval cathedrals, prehistoric villages, destroyed works of art. An app for detecting allusions in literary text is available.

Other digital humanities projects transform other kinds of data, collected earlier, into quickly understood images, much as, thirty years ago, scientists turned to artists to help make images out of otherwise incomprehensible supercomputer scientific data. Transforming data is only one modest part of what the digital humanities will be in the future, but you must start somewhere. People who resist this will find their arguments, and perhaps, as Edmond Campion worries, their work, made obsolete by the field's evolution.

[2] You can find the Homer Multitext Project at http://www.homermultitext.org/

Scale is another revelation of the digital humanities. An individual scholar or two can examine concepts or themes in thousands of books, not the tens once possible. In 2010, Stanford literary critic Franco Moretti began to urge his colleagues to try not close, but "distant" reading, computer-assisted reading of thousands of texts at a time. His Stanford Literary Lab has examined loudness in the 19th-century novel (Katsma, 2014) and the evolving language of World Bank Reports (Moretti & Pestre, 2015). That World Bank study, which showed a drift over 60 years toward more abstract and self-referential language, led to its chief economist, Paul Romer, demanding that its publications reduce their use of the word "and"—which led to the reduction of that economist's management duties (Schuessler, 2017).[3] Since retirement, Moretti has gone to Lausanne and is helping set up a new digital humanities program at ETH, the premier Swiss polytechnic.

Ted Underwood at the University of Illinois; David Bamman, University of California; and Sabrina Lee, University of Illinois used a machine learning algorithm to examine characters and authors in 104,000 novels. They noticed some unexpected trends: between 1800 and 1970, women decreased from 50 percent to 25 percent of authors of published novels, though that proportion later picked up: by 2000 they represented 40 percent. Women characters suffered a similar decline. Descriptors of women changed, too. (Eschner, 2018)

A literature post-doc at the University of Notre Dame, Dan Sinykin (2018) wrote an essay for the Perspective section of *The Washington Post,* which begins:

> I earned a PhD in literature the traditional way, reading a lot and

[3] Never mind. In 2018 Paul Romer won a Nobel Prize in Economics.

> reading carefully. By the end, though, I began to wonder at the provenance of the books I studied. What led them to me? What forces guided me to read one book and not another? Hoping to find out, I followed the money. In 1960, basically every U.S. publisher was independent, not owned by a greater entity. By 2000, 80 percent of trade books were published by six global conglomerates. What had the shift done to literature?

Making sense of a problem at that scale would require tracing trends and patterns across thousands of books, a feat beyond the capacities of a single human mind, but computational analysis offered a way. Sinykin's examination is underway.[4]

Most striking, many of these digital humanities studies are open. Anyone with the interest or the skills can participate. Projects are led by scholars but often gratefully crowd-sourced. This is very different from the humanities where I came of age, and a priestly caste ruled.

MIT, in announcing its new College of Computing, to begin in the fall semester of 2019, said its goal is to educate "the bilinguals," the people in fields like biology, chemistry, politics, history and linguistics who are also skilled in the techniques of modern computing that can be applied to their fields. "We're excited by the possibilities," Melissa Nobles, dean of MIT's School of Humanities, Arts and Social Sciences. "That's how the humanities are going to survive, not by running from the future but by embracing it" (Lohr, 2018).[5]

[4] Sinykin provides updates on his research via http://www.dansinykin.com/digital-humanities.html

[5] Actually, a College for Computing, but largely driven by rapid developments in AI. You may judge for yourself why this appeared in the business section, not the news or even the cultural section of the newspaper, even though the announcement stressed that this is a major intellectual turning point for the Massachusetts Institute of Technology.

Some in the humanities find the whole digital project worrying. They fear being disintermediated, or left out. They fear that scholarly or aesthetic judgments, which should be made by specialists, will be made by computer programs instead. They fear that the aesthetic encounter between humans and art will somehow disappear. I'm sanguine about the aesthetic encounter between humans and art, but the rest remains to be seen.

Whether the computer is mere instrument or has larger aims in the humanities is nowhere settled. Anne Burdick and her colleagues, in their book *Digital Humanities*, wrote:

> Digital Humanities . . . asks what it means to be a human being in the networked information age and to participate in fluid communities of practice, asking and answering research questions that cannot be reduced to a single genre, medium, discipline, or institution . . . . It is a gobal, trans-historical, and transmedia approach to knowledge and meaning-making.

Digital Humanities is itself a model, a "collaboratively crafted work," each of the five authors originating and editing the final product, trading the manuscript back and forth electronically. The authors claim this for the digital humanities in general: they are "conspicuously collaborative and generative" (Burdick et al., 2012).

So the digital humanities thrive. Stanford claims involvement in digital humanities (although sometimes under different names) since at least the late 1980s, encompassing literature, music, history, anthropology, and much more. At Harvard, the introductory computing course—for majors and nonmajors alike—fills venerable Sanders Theater every semester. At Columbia, a course called Computing in Context aims to teach humanities students programming and computer science logic within the context of three of their own disciplines, English, history, or economics. General lectures by a computer scientist occur twice weekly, and professors of English, history, and economics conduct the discussion sections.

A provocative collection of essays called *Defining Digital Humanities* includes a quote from Willard McCarty, a professor of humanities at King's College London and a fellow of the Royal Anthropological Institute (and known fondly as the Obi-Wan Kenobi of the digital humanities). McCarty says, "I celebrate computing as one of our most potent speculative instruments, for its enabling in competent hands to force us all to rethink what we trusted that we knew" (Terras, Nyhan, &Vanhoutte, 2013, p. 5).

Rethinking what we trusted that we knew: this is the breathtaking challenge of the present-day humanities.

But all this has been about the encounter between the humanities and the computer. Where does AI fit in? One place is AI's great inferential abilities, reading and drawing conclusions from all those texts. But that's only the beginning.

3.

Defining Digital Humanities includes the essay "What Is Humanities Computing and What Is Not?" first published ten years earlier by John Unsworth, now University Librarian and Dean of Libraries at the University of Virginia. In the essay, Unsworth considers what it means to reason intelligently and how we make inferences based on what we know, questions that have been central to the humanities.

Unsworth quotes extensively from a key 1993 AI paper, "What Is Knowledge Representation?" by three MIT professors of computer science: Randall Davis (a former PhD student of Ed Feigenbaum's), Howard Shrobe, and Peter Szolovits. In quoting the paper, Unsworth shows that artificial intelligence has brought these questions from the humanities into sharper focus because AI has had to deal with the same questions more exigently and precisely.

The humanities are very much about the representation of knowledge, Unsworth says, and it's time humanities scholars acknowledged that: "In some form, the semantic web is our future, and it will require formal representations of the human record. Those representations—ontologies, schemas, knowledge representations, call them what you will—should be produced by people trained in the humanities." I'd reply that representation has not been neglected in traditional humanities studies, but it was mostly about identifying genres and styles, or naming movements: allegory, the novel, irony, Modernism, post-Impressionism, identifications that were often vague and elastic.

Why does it matter? It matters, Unsworth says, because we are entering a new world. To navigate this new world, we need formal representations, which must be computable, because the computer mediates our access to this new world. Finally, those formal representations must be produced first-hand by those who know the terrain. Yes, he concedes, these will be maps, and maps are always schematic and simplified, but that's what makes them useful.

Ontology—the nature of being, or existence—is seldom addressed outside a philosophy classroom, but it emerges as deeply important in computational models, because certain ontological questions must be answered: What exists? What categories can we sort existing things into? What's universal and what's specific? How can a body of knowledge maintain its consistency? AI has sharpened these questions for its own purposes and is useful in showing the humanities how to ask and answer such questions in computational models.

Ten years after the essay's first publication, Unsworth added a commentary to his paper in praise of AI as a humanities model:

> It seems important to establish that 'humanities computing' is not just an instrumental term, with the focus on using the computer, but an intellectual activity in its own right. Or maybe not exactly in its own right: as an intellectual activity it appears to require validation in terms of another field of inquiry (artificial intelligence). (Terras, Nyhan, &Vanhoutte, 2013)

Well, now.

4.

There's much to celebrate in these new efforts and what they represent: a recombination of scholarly pursuits from fields across the spectrum. As with all recombinations, new things will appear: new tools, new points of view, new knowledge.

For instance, for more than a decade, David Blei, a Columbia University computer scientist, has developed LDA (latent Dirichlet allocation), a powerful statistical tool for discovering and exploiting the hidden thematic structure in large archives of text. LDA aims to capture the intuitive belief that documents exhibit multiple topics, and each document in a large collection can be situated. Contrary to human-like approaches, LDA assumes that text can be considered a "bag of words"—the word order in a given document doesn't matter, and neither does the document order in a collection. In short, LDA has no semantic understanding.

LDA has proved to be deeply helpful in teasing out hidden themes in large collections of documents, from survey data to population genetics. But the applications that amaze Blei most are in the digital humanities: historians and English professors and folklore scholars, who want to find patterns—recurring themes—that they wouldn't otherwise have noticed (Krakovsky, 2014). For example, Blei says:

> Matt Connelly, a historian at Columbia, studies the history of diplomacy, and he has a dataset of all the cables sent between different diplomatic stations in the '70s. He could use LDA to analyze it, or he could say 'I know something about this,' and I can sit down with him and build a topic model based on what he knows about the data. Finding the hidden structure that this data exhibits is a computational problem called inference—the problem of computing the conditional distribution of the hidden variables, given the observations. (Krakovsky, 2014)

Some humanities scholars have seized LDA as another of many lenses to discover patterns that wouldn't be found by close reading and moreover, to make tractable uncovering the themes in thousands, perhaps millions, of documents, a job beyond any single human head, or team of human heads.

Inference was the great task of early expert systems of the 1980s, but those systems were painstakingly handcrafted. Nowadays, far more complex and sophisticated algorithms, infinitely faster processing, bigger memories, and orders of magnitude more available data have changed the game, an example of quantity that transforms the quality of research. Human experts like the historian Connelly can certainly help prune the search space, the way experts once contributed heuristics to simpler intelligent programs, but the machines obviate year-long trips to obscure archives and libraries (a pity!) at the same time they tease out themes in our own productions that would be difficult, perhaps impossible, for us to see otherwise. They do it in ways unlike how we think. It gives me delighted pause.

On the other hand, Blei pointed out to me in a talk we had, in the old days, you were limited by your data in a good way. Your models were parsimonious. Massive data sets change the game. Models are more inclusive, perhaps, but also more unruly, vexatious, and by virtue of their provenance, harder to validate. Machine learning still struggles with data that can't be quantified, Blei said, and often shoehorns what was once a problem of prediction, ML's earliest task, into problems that have little to do with prediction. A human intelligence must still look at the results of LDA, and decide what they signify (as humans must supply labels to images for ML).

Blei told me he loves to go to digital humanities workshops and discover what scholars are up to and what they need. His findings push him, and his students, to do research that makes better and more useful tools for research. "It's a wonderful feedback loop for us."

No mere trend, Wendell Piez at the University of Illinois assures us, digital humanities *are* humanities in the digital age. Strange as this all may seem, he argues, we've been here before:

> Digital humanities represents nothing so much as the *humanistic* movement that instigated the European Renaissance, which was concerned not only with the revival of Classical scholarship in its time but also with the development and application of high technology [printing] to learning and its dissemination. Scholar-technologists, such as Nicolas Jenson and Aldus Manutius designed typefaces and scholarly apparatus, founded publishing houses, and invented the modern critical edition. In doing so, they pioneered the forms of knowledge that

> academics work within to this day. . . (Terras, Nyhan & Vanhoutte, 2013) [6]

Digital humanities are for the generations of students, eventually future scholars, for whom computers are not a specialized tool, but "part of the tissue of the world," writes Julia Flanders, head of the Digital Scholarship Group at Northeastern University. Moreover, because digital storage is so cheap (cheaper than making decisions), neglected and otherwise overlooked works are digitized and accessible, aggregated "into noticeable piles, so minority literatures, non-canonical literary works are now visible." (Terras, Nyhan, &Vanhoutte, 2013) Texts outside the canon—the largest body by far of human texts—are what Thomas Leonard, former University Librarian at Berkeley once told me he calls "the great unread." Until now, they've been all but inaccessible, whether because they were filtered by publishers who calculated a book was inappropriate for commercial publication (e.g., because the book's topic was too geographically localized, or its language had insufficient readers to be profitable). Self-publication allows innovation that can evade the conservatism of commercial publishers and editors, but can also lead to prose that in any language is downright unreadable.

A glance at any of the new books or scholarly journals devoted to the digital humanities confirms that a new field has lively differences among its present practitioners, to say nothing of what its detractors and scoffers say. "But the intellectual outcomes will not be judged by their power or speed, but by the same criteria used in humanities scholarship all along: does it make us think? does it make us keep thinking?" Flanders adds.

[6] Quotations from Piez and Flanders appear in Terras, Nyhan, and Vanhoutte, 2013. The quotation from Thomas Leonard is from an author's interview.

But these examples are the humanities seizing and employing digital tools of many kinds from information processing. AI professionals, for their part, implore the humanities to contribute to the AI enterprise, to make the project more successful and ethical, and to improve it in every way for human benefit (AI Index, 2017).

5.

We come back to C. P. Snow and his Two Cultures challenge. We might go further back, to Thomas Henry Huxley three-quarters through the 19th century (a naturalist known as Darwin's bulldog) who claimed the quarrel actually began in the 18th century between partisans of ancient literature versus those of modern literature and shifted in the 19th century to humanities versus the sciences. Huxley observed this when speaking at the inauguration of a science college, eventually to be the University of Birmingham. Snow merely revived the theme in the 1950s. My college freshman reading included Huxley's speech, but it fell on blind eyes then (Huxley, 1875). Now I see the sciences and, miraculous to say, engineering (that Cinderella of the universities, eternally sweeping up ashes, its neglect a result of 19th-century Romanticism, which defined anything practical as negligible) both occupy the center of 21st century intellectual ferment.[7]

So the digital humanities borrow much of AI's intellectual

[7] Huxley on the topic: "How often have we not been told that the study of physical science is incompetent to confer culture; that it touches none of the higher problems of life; and what is worse, that the continual devotion to scientific studies tends to generate a narrow and bigoted belief in the applicability of scientific methods to the search after truth of all kinds" (Huxley, 1875. Science and Culture, and Other Essays. Project Gutenberg, www.gutenberg.org/ebooks/52344) And so on, with bracing Victorian confidence. You could even argue that the division goes back to the Greeks, who distinguished between episteme, theoretical knowledge, and tekhne, tools, methods for achieving results.

endowment. In AI, a Janus-faced entity emerges: its one face tools, its other, mirrors. The tools are to excavate more thoroughly and construct more precisely and inclusively who humans are, were, and might be. The mirrors reflect us and our everlasting preoccupation: it's all about us. AI isn't alien—though it could one day be, which would be a third face. It's what we want to know and what we care about heightened and made more precise. How else could AI be, given our wired-in self-absorption, this remarkable adaptation that has helped us perform for eons the exacting dance between cooperation and competition among individuals and among groups?

A good thing, I'd say, for as we'll see in a subsequent chapter, we'll need every resource we have to meet what philosopher Nick Bostrom calls the essential task of the century.

Part Seven: And Wherefore Was It Glorious?

And wherefore was it glorious? Not because the way was smooth and placid as a southern sea, but because it was full of dangers and terror, because at every new incident your fortitude was to be called forth and your courage exhibited, because danger and death surrounded it, and these you were to brave and overcome. For this was it a glorious, for this was it an honourable undertaking. You were hereafter to be hailed as the benefactors of your species, your names adored as belonging to brave men who encountered death for honour and the benefit of mankind.

—Mary Shelley, *Frankenstein*

Elegies

1.

All flesh is as grass, and many of my teachers and mentors in artificial intelligence have died. The voices of the dead are said to be what we forget about them first, but I hear the distinct timbres, the laughter, the rhythms.

Allen Newell's death was too early—he died of cancer in his sixties, and although Joe and I were able to say goodbye to him, our loss was deeply personal, enduring, and made all the more melancholy by the work that remained undone.

His work in the last years of his life was enviably ambitious. From the 1950s Logic Theorist, designed to simulate only "a small but significant subset of human thinking," he eventually proposed and brought forth a program called Soar. Soar adapted the model Newell had first presented in 1980 of levels of intelligence in the computer from the zero-one level all the way up to the highest, what he called the knowledge level. Now he proposed Soar as a unified theory of *human* cognition whose details are in his 1987 William James Lectures (Newell, 1990). A multilayered, asynchronous model of human cognition seized scientific imaginations, and the late 1980s and early '90s saw a spate of such programs, including stunning breakthroughs in a subset of machine learning, known as deep learning. Deep learning, as I've said, was invented by Geoffrey Hinton and his colleagues, Yann LeCunn and Yoshua Bengio about the time Newell was giving his James lectures, but needed nearly another three decades until computing technology became powerful enough to implement it so fruitfully. Hinton, along with LeCun and Bengio, won the 2019 Turing Award for their development of deep learning.

Newell died on July 19, 1992, aged 65. The days of Allen's dying and death must have been unendurable for Noël. Yet endure she did. Soon I began to hear that Noël had started traveling. For those of us who'd known how fragile she was in Allen's lifetime, this seemed inconceivable: we swapped the stories, slightly disbelieving. But there were the facts—to Europe, to Asia. I saw her briefly at a Pittsburgh dinner sometime in the late 1990s. She'd just returned from Vietnam, where she'd crawled through the tunnels used by the Vietcong during that tragic war.

Allen's devotion to science, his exacting standards for himself, and his disdain for those who didn't measure up might have been more of a burden to live with than we'd realized. Certainly the facts were that, at Newell's death, Noël was released and took wing. For all the lugubrious yearnings she and I shared in the 1970s, she didn't go back to San Francisco. If anyone in her adoptive family remained, what did they mean to her? Allen's parents were gone. She'd made her life in Pittsburgh after all. Unlike the stoic Dorothea Simon, who fled to California after Herb's death, to spend her remaining days with her sister there, Noël stayed, tended the flame of Allen's legacy, and lived a life of her own for the first time. She was admired for her courage—in her early eighties, she'd slipped a few months earlier on winter ice, broken a hip, but came to a talk I gave at CMU's School of Computer Science, stepping smartly if slowly with her cane, alert and amused, fully self-possessed. I saw her again in 2015 at the fiftieth anniversary celebration of the Carnegie Mellon computer science department, and she was lively and mobile, despite some eye surgery; she calculated for me that her great travels had gone on for seventeen years after Allen's death. Once more in 2018, at age 90: "Don't tell me how good I look!" she hissed. But she did look good.

2.

After I left Pittsburgh, Herb Simon and I met from time to time, always warmly. When Columbia University dedicated its new computer science building in 1983, he received an honorary degree at the occasion. We sat together through the opening talk by Columbia's provost, who spoke nearly verbatim remarks I'd prepared for him. Simon leaned over and whispered to me: "Do you know whose ideas those are?" I laughed. Sure, Simon, recycled by McCorduck. Simon corrected me: "Alan Perlis, Allen Newell, *and* Herb Simon." The following day, *The New York Times* ran a classic picture of Joe putting the doctoral hood on Simon's shoulders, with Columbia's president standing by. The cutesy story that accompanied it (by a new reporter called Maureen Dowd) seemed to me yet another effort to put those unsettling machines in their place (1983).

A few years later, Simon was my dinner partner at a New York Academy of Sciences affair. Also at the table were several others, including Donald Knuth, a giant in the history of algorithms, whose volumes are known simply as "Knuth." He had loosened up with dinner wine, and said pointedly to me, "I suppose I shouldn't ask this, but are you ever going to do Herb's biography?" Simon said: "She's waiting til I've done enough to fill a book."

Eventually, Simon wrote his own memoir and sent me a draft. It was lengthy and candid. When I was thinking of doing his biography, he'd warned me away from writing anything personal, but to my surprise, he confessed that many more of his relationships with women might not have been platonic if he hadn't been fearful of rejection. The manuscript described a mind-affair he had not too many years after he was married, deftly handled, honest, just the kind of thing I couldn't have put into a biography, but it fills out the portrait. Although I worried the manuscript must be shortened, I loved its luminous good cheer.

Perhaps the last time I saw Herb Simon was at the 1990 25th anniversary celebration of Carnegie Mellon's computer science department. It was a warm and lovely few days, celebrating the past, yet mortality was haunting us, with Allen Newell dying in his Squirrel Hill home. Ed Feigenbaum and Penny Nii were there, and we all savored each other's friendship, took snapshots "like Japanese tourists," I wrote in my journal after a lunch together. Along with his close colleagues, like Pat Langley, Simon's work now was about simulating the process of scientific discovery. Their Bacon program had rediscovered Kepler's third law and Ohm's law.

Bacon would be a precursor to programs that work not in the history of science but its future. Descendants of that program are now at the frontiers of science. Yolanda Gil and her colleagues at the University of Southern California write that these programs can

> radically transform the practice of scientific discovery. Such systems are showing an increasing ability to automate scientific data analysis

> and discovery processes, can search systematically and correctly through hypothesis spaces to ensure best results, can autonomously discover complex patterns in data, and can reliably apply small-scale scientific processes consistently and transparently so that they can be easily reproduced. (Gil, Greaves, Hendler, & Hirsh, 2014)

Not bad. Moreover, Gil and her colleagues write, "AI-based systems that can represent hypotheses, reason with models of the data, and design hypothesis-driven data collection techniques can reduce the error-prone human bottleneck in scientific discovery." Even better.

These new techniques aren't limited to text: they analyze nontextual sources, such as online images, videos, and numerical data. "The world faces deep problems that challenge traditional methodologies and ideologies," Gil and her colleagues continue. "These challenges will require the best brains on our planet. In the modern world, the best brains are a combination of humans and intelligent computers, able to surpass the capabilities of either one alone" The Defense Advanced Research Projects Agency (DARPA), one of the original sponsors of AI research, has begun automating some research this way, still adhering to its mission to invent revolutionary technology.

Joe and I often tried to coax Simon to visit the Santa Fe Institute. The Institute's core research focuses on the sciences of complexity, where complexity arises from simplicity. All of the Institute's original scientists understood their debt to Simon's ideas about complexity; everyone quoted *The Sciences of the Artificial*, where the ideas had been laid out. Simon would've been warmly welcomed. Maybe he was already feeling too old to travel just for lionization, and, as he once said about China, he could learn more in the University of Pittsburgh library than he could by visiting (although he loved traveling to China and often did). We failed to bring him to Santa Fe.

My journal is oddly silent about Simon's death in February 2001, aged 84. I didn't go to his memorial service, although Joe did. Did I need to meet a class I was teaching at Columbia in writing about science? I regret missing the memorial service, but more, I miss Herb Simon. He's still alive to me in his intellectual acuity and his capacious ability to synthesize and make connections between deeply different fields, dissolving the outerwear of disciplines to find their commonalities. He's still alive to me in his joyous laughter.

In November 2013, Carnegie Mellon announced the launch of the Simon Initiative. Named in Herbert Simon's honor, the Simon Initiative is a cross-disciplinary initiative in which learning science impacts engineering education and vice versa. As the world's largest database on student learning, it examines the uses of technology in the classroom, identifying best practices, helping teachers to teach, accelerating innovation and scaling through start-up companies (a specialty at Carnegie Mellon), and improving student educational experience. It follows from Simon's long interest in the cognitive sciences involved in teaching and learning. Its scope is now international; anyone can contribute to or use it. Dan Siewiorek says the Initiative has two aspects: the deeper science underneath teaching and learning, and a higher vision of those that goes well beyond the much-hyped MOOCs (massive online open courses).

Also on the Carnegie campus, Newell-Simon Hall honors them both. Housed in Newell-Simon Hall is an extraordinary if gawky-looking robot called Herb, a prototype household assistive robot that can find and manipulate objects in the visually confusing environment of the ordinary household.

One of the most maddening things about Simon's legacy is how fundamental his ideas became in so many fields that they began to be counted as derived from God. Ed Feigenbaum and I sent outraged messages to each other each time Daniel Kahneman was described as "the first behavioral economist to win a Nobel Prize." No, Herb Simon, who destroyed the myth of rational man in economics, was the first.

3.

Over the years, John McCarthy and I ran into each other from time to time at meetings and other events, and he would entertain me, as always, with new ideas and stories to illustrate how knowing the science and doing the calculations mattered. My favorite do-the-calculation example is his Magic Doctor. A young doctor is inexplicably gifted with the ability to heal anybody he or she touches. McCarthy proposed various outcomes—that the poor doctor soon drops from fatigue, is sequestered by the wicked and economically threatened medical establishment, or is assassinated by a religious lunatic who believes suffering is the proper fate of sinful humankind. But McCarthy shows how, by doing the arithmetic, the young doctor can in fact heal everybody on earth afflicted with disease in a few hours each day. (The details are in my book, *The Universal Machine.*)

"It's what I call the literary problem," McCarthy said to me more than once. "You can't make stories out of things working well. You need conflict, failure, drama, to tell a story. That's why most science fiction is dystopian. It wouldn't work as narrative if everyone lived happily ever after."[1]

[1] Alex Garland's 2015 film, Ex Machina, is an entertaining example. Saturated in literary precedent (from Pygmalion to Frankenstein to R.U.R., with a nod to Bluebeard's

Ed Feigenbaum threw himself a sixtieth birthday party at a Silicon Valley funhouse, whose main entertainment was for the guests to climb into a flight simulator and pilot whatever kind of aircraft (or maybe spacecraft) they wanted to try. So woozy in the demo that I never made it to the simulator, I staggered out to look for other diversions. With great good luck, I found John McCarthy leaning against a wall, uninterested in the electronic entertainment, I guessed, because he was already a licensed pilot and had done the real thing.

We fell to talking, and this night McCarthy was especially animated. Did I know his daughter Susan's work? She'd had a best seller as the coauthor of *When Elephants Weep*. I was astonished. "Sumac is your daughter?" I knew Susan McCarthy from the online forum, The WELL; we belonged to a small group of women who chatted online with each other almost daily. John's daughter? No, I certainly hadn't known. I'd followed the adventures of Susan's children as they graduated from high school, moved on to college, moved out into the world, never knowing I was also hearing about John McCarthy's grandchildren.

It was touching to see his pride in her, not only in how well she was doing as a writer, but how she'd dedicated her life to watching and caring for animals in the wild. Susan would continue to write sharp-eyed, witty pieces about all sorts of wild animals (from bugs to crustaceans to blue whales) on her blog and published another book, *Becoming a Tiger*, about how baby animals learned to be grownups of their species. Later, she'd coauthor an acidic but funny blog called SorryWatch, which calls out public figures for their sleazy non-apologies.

Castle), it has a well-marked hero, villain, and—well, just what is the robotic woman? You can find precedent for her character in each of those works.

When Susan McCarthy told me she'd be coming with her father to New York City for *The New Yorker* Festival in September 2002, I insisted they be my guests for lunch. It was wonderful to see them both together, teasing each other, taking pleasure in each other's company. It might be at this lunch that John McCarthy expanded on his quarrel with literature:

> When stories take up the theme of technology, especially AI, it's always dystopian. I suppose in a story you need conflict, you need an *us* the reader can identify with, and a *them* the reader can root against. The *them* to root against is always some technology. In stories, AI is always out to get us, and we must outwit it. Given the conventions of stories, I don't see how that can be fixed. But in real life, it's simply not so. Technology is mixed, but on the whole, it's been a tremendous benefit to the human race. There's no reason to think AI will be any different.

After lunch, I took them out to hail a taxi. I insisted that John McCarthy give me a goodbye hug. Awkwardly, he did. It was the last time I saw him before he died in 2011. Susan McCarthy swears that somewhere in the house are her father's notes for how technology came to, and improved, Tolkien's Shire.[2]

4.

In the fall of 2013, many years after my original interviews with Marvin Minsky, I sat in the same room where I'd once heard him play his music and watched him try to fix his wife's CPR dummy. Gloria Rudisch Minsky was with us, and I reminded them both of that moment. Gloria remembered at once and began to laugh merrily. "That dummy never did work very well!" she exclaimed. "But the

[2] When I heard some dozen years later that The New Yorker was experimenting with an AI to deal with the volume of entries it received for its cartoon caption competition at the end of each issue, I wished John alive so I could laugh about this with him.

dummies are much better now." As she'd get up to go toward the kitchen, moving from table to table to support herself (she walked with difficulty now), sometimes her hands would reach for the couch where her husband was sitting, and their fingers would touch for a moment, reassuring each other silently, lovingly.

At age 87, Minsky looked astonishingly unchanged. He was trim, upright, his face unlined, his entire cranium radiating intelligence, as it always had.[3] He told me he was still composing music. Serious health problems had slowed him, but with luck (and once, thanks to his wife's fast diagnosis), he'd weathered them well. We sat in the same crowded room, every surface, table, floor, covered with odds and ends, such as a little Christmas theater, toy trucks still in their boxes, sheet music, a giant wrench and screw over the fireplace (each item scrupulously dusted, which signaled it was very much an intentional collection), the harmonium and the piano still in place along with assorted other keyboards too, books stuffed into shelves and lining the staircase. We drank tea, ate cookies, and found much to laugh about.

Finally, I asked Marvin where he'd like to see artificial intelligence go in the future. He didn't answer immediately. At last he said, "I'd like to see it step in where humans fail."

Minsky was still active at MIT when I saw him then, planning to teach the AI course the following semester. "What will you say?" I asked casually. He responded, "Oh, I'll probably just say, 'Any questions?'"

[3] Patrick Winston likes to tell this story: Danny Hillis once asked him if he ever had the experience of telling somebody a new idea, only to have his listener misunderstand, and in his misunderstanding, make it a better idea. "Almost every time I talk to Marvin," Winston replied.

He died on January 24, 2016, of a cerebral hemmorhage. Susan McCarthy, John's daughter, doing research then in Antarctica, had called him and Gloria a few weeks earlier. They were thrilled to be called from the Antarctic, she said, Marvin his usual self, but sounding very weak. He was the last living of the four founding fathers of artificial intelligence, and everyone who knew him knew a giant intellect, an inspiring teacher and mentor.

5.

Why did none of these four share the fevered fears of later scientists, like Stephen Hawking, or entrepreneurs, like Elon Musk? One answer is *ars longa, vita breva*, and success seemed very far off. Better to let the problems be met by people who actually needed to grapple with them than lay down hypothetical rules that would be overtaken by reality and time.

But in addition, the founding fathers were all realists. As John McCarthy had often said, technology is mixed, but on the whole, it's been an enormous benefit to the human race. Why should AI be any different? Yes, it would require many adjustments, some of them major—imagine not having to work at disagreeable, boring jobs just to keep body and soul together—and prudent governments would eventually understand how economic security for citizens was not only needed, but easy to supply. People could then take on tasks that might give them satisfaction, for humans do not like to be idle or without purpose.[4]

Current researchers already aim to build systems that extend, amplify,

[4] Given the prevailing attitudes of our time, say sages of the second decade of the 21st century, a basic minimum income is a nonstarter. But prevailing public attitudes can change very quickly: examples include attitudes toward gay marriage or sexual harassment.

and provide functional substitutes for human cognitive abilities. "A principal goal of applied AI is and should be to create cognitive orthoses that can amplify and extend our cognitive abilities. That is now and near; a computational Golem is not" (Ford, Hayes, Glymour,& Allen, 2015).[5]

These orthotics will assist the normally aging, or others with small cognitive disabilities. AI already helps operate exoskeletons, devices that allow disabled people to stand upright, walk, and use their arms in easy, intuitive ways. Rehabilitation robots can physically support and guide patients' limbs during motor therapy, but to do that successfully requires sophisticated AI. In partnership with other disciplines, AI is poised to transform the experience of learning; in both formal and informal settings, classrooms of the future will be places "to achieve challenges together rather than . . . places where teachers teach and students listen and do problem sets" says Janet Kolodner (2015). Her view is far more capacious than the old teaching machine environment, and emphasizes especially the need for collaboration across disciplines.

6.

My best teacher in science, as in so many things, was my husband, Joseph F. Traub, who died suddenly in the late summer of 2015. I'd been a humanities student, but Joe continued the education Ed Feigenbaum had begun and taught me to think like a scientist, too. Why? Where's the evidence? Is that what the evidence really says, or is there a different way of looking at it? Is it sufficient evidence? Validated evidence? What theory can we abstract from this? Suppose

[5] The entire Winter 2015 issue of AI Magazine (in which this quote appears) is devoted to AI for rather than instead of people and is a bracing corrective to the fevered fears of the last few years in the popular media.

instead, we. . . What if? I *wonder*. . . He set a great example—and I was open to it—of working hard and playing hard.

He'd overseen the transformation of the Carnegie Mellon computer science department in the 1970s from ten faculty to fifty professors and researchers before he left for Columbia, and his intellectual legacy there was commemorated in 2015 by Carnegie Mellon with a chair named in his honor. His *New York Times* obituary mentioned how he'd been recruited to his alma mater, Columbia University, to bring computing to one of the great Ivy League universities. In his oral history, in the archives of the Computer History Museum in Silicon Valley, he said that his challenge was "to convince one of the great arts and sciences universities in the United States that computer science was really central" (Raghavan, 2011). You've seen from these pages that this wasn't simple.

In some ways we faced a parallel task, Joe with a great but decidedly backward university when it came to computing, and mine with the literary leaders of my generation. But he began the task and lived to see it carried out by an energetic young computer science faculty. At the Columbia University memorial service in his honor, a common theme emerged: he was a sensitive mentor to young people. Some of them were middle aged and in mid-career now, but each spoke of how Joe had guided, advised, and nurtured them and, in one case, beat on the Columbia bureaucracy to allow Kathy McKeown, a woman with two little babies, to be allowed some slack leading up to tenure so she could have both babies and a career, a revolutionary idea in the early 1980s.

As he had at Carnegie Mellon, he pursued his own scientific career in parallel with building a computer science department at Columbia, doing research and publishing until the end of his life.

Joe was equally active in public service: the founding chair of the Computer Science and Telecommunications Board of The National Academies of Sciences, Engineering, and Medicine, the country's leading advisory group on science and technology, which he served twice as chair. After that, he moved on to serve on a board of the National Research Council.

Although his own research was distant from it, Joe was an ardent supporter of AI research. Allen Newell and Herbert Simon introduced him to AI, and he figured that the smartest people he'd ever known must be on to something. When he began hiring at Columbia, he looked first not for specialists in his own field, but for AI people—he believed they'd be the intellectual leaders in a new department. He propped me up when I wanted to collapse from the frustration and difficulties of writing *Machines Who Think*; he shared my enthusiasm for later work; he was very glad I was writing the human side of the story in this book.

In the meantime—I take a breath to say it—he pursued other passions: he loved international travel; took cooking lessons in Pittsburgh from the man who would eventually head the Culinary Institute of America. When we moved to New York, he took several courses at Juilliard—and then convinced me I should join him—to learn more about the music we both loved. He signed us up for courses at the Museum of Modern Art.

He assembled a wonderful collection of early computational instruments. We went to European flea markets at dawn and the annual auction of "office machines" in Cologne, or Joe asked friends, then friends of friends, who might know how he could get his hands on a particular instrument. This eventually included two Enigma machines, a three-rotor and a four-rotor (all now on exhibit at Carnegie Mellon's Hunt Library).

In the summer of 2012, we traveled to Alsace, mainly to eat and drink. At the town of Colmar, we expected to see a celebratory exhibit of early arithmometers, one of the first widely distributed digital calculating machines, manufactured by Thomas de Colmar in the early 19th century. We owned a couple of these handsome instruments. But curators at the municipal museum stared at us blankly. Somebody allowed as how a few arithmometers might be stashed in a warehouse outside town, but on exhibit? Why would that be interesting? The beginning of the digital age, Joe explained patiently. You should honor this distinguished son of Colmar.

In the last ten years of his life, Joe returned to his first love, physics, partly because he thought his research might have applications to quantum computing and partly because he loved to know. He was thrilled by all the new physics discovered since he'd been a graduate student at Columbia so many decades earlier.

He loved the outdoors, especially the mountains, which he climbed and skied in when he was younger and hiked in until three days before his death.

He loved me. I used to tease him that I was the pampered darling of a doting husband, and he agreed unashamedly. It was an intimate marriage of nearly half a century, emotionally and intellectually, that gave us deep pleasure, joy, and strength.

The Male Gaze

1.

At the end of his book *Chance and Necessity*, Jacques Monod, a molecular biologist and Nobel laureate, asks humanity to put its faith in art and science for human salvation. For a long time, I could see why science might save us collectively. I was less confident about art. It's cliché to say that Nazi commandants loved Brahms and Wagner and still behaved inhumanly, but it's true and hasn't yet been satisfactorily answered on behalf of art. But if art only enriches the lives of those individuals whom science saves, then perhaps that's enough to ask of art.

As a humanist, I soon understood that AI was a human enterprise. From its beginning, it was, and still is, two-pronged: first, to model and therefore better understand human cognition, and second, to apply better-than-human cognition to various problems humans want to solve. I was curious about AI's history, its prospects, and the people who were making it happen. The computer, I thought, was a tool, human thought made manifest (even before I tangled with Ashley Montagu on that). As you see from these pages, this saga, I was innocently surprised to discover that other humanists didn't view AI or computing in general the same way. They saw my curiosity as "selling out to the machines," as scheming to "exchange machines for humans," and other fearful nonsense.

"As a substitute for humans?" Oliver Selfridge once mused to me. "How? Will the machine substitute me for myself? Of course not." But over my lifetime, most humanists just dismissed AI.

In 1989, a simple question on a publisher's publicity questionnaire about "how I came to write this book"—*Aaron's Code*—stopped me. It wasn't just one book; it was my life's *fata morgana*. Why artificial intelligence? What drew me and never let me go?

July 21, 1989:

> *Simple question, but the answer is by no means simple. That early on, I recognized how significant AI was, and wanted to tell the world? Paltry themes bore me. I have no desire to paint teacups. I want to run with the big dogs, but in my own way. Seize a bone you can hardly get your jaw around, then hang on for dear life.*
>
> *This long romance with AI? The appeal to my rational, cool, cerebral self, yes. I admire in it what I admire in myself (like everybody else doing AI). Yet surely I wanted to winkle out the passion in these people too. I knew my own.*

The easy answers: How much I admired the people engaged in the research—astonishing visionaries, pursuing with acute natural intelligence what had been a human dream forever, but never before possible. If they succeeded, they'd change the world. (And so they have.)

Herb Simon, Allen Newell, John McCarthy, Marvin Minsky, Seymour Papert, Raj Reddy, Ed Feigenbaum, Harold Cohen, Lotfi Zadeh, Tomaso Poggio, Patrick Winston, and so many others, all of them at extremes, stretching the mind—their own, mine, the world's, and newly created artificial minds—in unprecedented directions. I was enchanted by their vision.

Some of them, like Newell and Simon, like Minsky, McCarthy, Feigenbaum, Poggio and Winston, pursued deep scientific goals. Some of them, like Reddy and Papert, wanted to free ordinary people, whether in poverty or just in poor schools, from wretched circumstances. Some of them, like Harold Cohen, had an artistic goal to pursue. For all of them, their goals were demanding, yet high-spirited fun. This young woman, depressed by the pessimism of post-World War II literature, loved the optimism, the excitement, and the ambitions of early AI research. All these were answers to that question.

Did AI represent an attractive transgression for me? A stand-in for being a Merry Prankster? Hardly. For most people in the coming decades, my romance with AI was so absurd it wasn't even transgressive. To them it was just—incomprehensible. Wacky. Preposterous. Likewise, my serene confidence in it.

Yet in conversation with Harold Cohen, I once blurted out that I thought if intelligence could be shown to exist outside the human cranium—outside the *male* human cranium—then all the stupidity I'd endured, the conventional wisdom from every side ("women mustn't be too smart; they won't get a husband," "graduate school isn't for you, you'll only get married and have babies," "the most valuable asset you bring to this organization is your typing speed") was exactly that, stupidity. In the 1970s, when the second wave of feminism began to free me and millions of other women, we finally saw those platitudes for the old-fashioned superstition and self-serving patriarchy they were. Intelligence exhibited by a machine made nonsense of the universally sanctified superiority of male thought. In my enchantment with AI, I could've been sticking out my tongue at all that.

As a corollary, I naïvely imagined that AI would be neutral intelligent behavior without sexism or other bigotries, that somehow symbolic reasoning and algorithms descended from some Platonic Ideal Heaven, free of traditional earthly biases. I wasn't alone in such a hope. Algorithmic decision-making in loan approvals, welfare benefits, college admissions, employment, and what social media shows its participants seemed to offer mathematical detachment and lack of bias.

But this neutrality was a subconscious hope for AI I kept even from myself for a very long time. How wrong I was. Algorithms arise from the soil where they germinate: they embody the unexamined assumptions of their programmers, a majority of whom are white or Asian men. These are the people who label images (in the beginning) and assign the weights of importance to one fact or another (in the beginning). Moreover, the data sets these already compromised algorithms work over are tens of thousands, often millions, of human decisions, and as those decisions exhibit bigotries, so does the program.[1] Silicon Valley has up to now shown itself feudally backward about shaping more inclusive versions of a social culture, so we can't be astonished that its products reflect that sorry state

[1] See, for example, Weapons of Math Destruction by Cathy O'Neil (Penguin Random House), Life in Code: A Personal History of Technology by Ellen Ullman (Farrar, Straus and Giroux), and Plain Text: The Poetics of Computation by Dennis Tenen (Stanford University Press).

of affairs. We can't be astonished that parts of AI might be equally biased.[2]

For a sorry example, in October 2018, Google employees were outraged by a *New York Times* report that scores of their colleagues, found to be sexual harassers, had been quietly let go, many with generous exit packages, including Andy Rubin, a creator of the Android mobile software, who left with $90 million (Wakabayashi & Brenner, 2018). Some 20,000 Googlers worldwide, female and male, staged a temporary walkout on November 1, protesting their firm's behavior and demanding more transparent handling of sexual harassment, and greater efforts on behalf of diversity. Google's president, Sundar Pichai, publicly apologized and promised that the firm would do better (Wakabayashi, Griffith, Tsang, & Conger, 2018).[3] But two of the protest's female organizers later claimed they'd suffered retaliation from Google.

[2] For a good, if brief, survey of sexism in Silicon Valley, see "Letter from Silicon Valley: The tech industry's gender-discrimination problem" by Sheelah Kolhatkar (The New Yorker, November 20, 2017). The Valley's ageism and racism is no better. See also "Amazon's facial recognition wrongly identifies 28 lawmakers, ACLU says" by Natasha Singer (The New York Times, July 26, 2018) and "How white engineers built racist code—and why it's dangerous for black people" by Ali Breland (The Guardian Weekly, December, 4, 2017), and too many more such articles. A particularly detailed and damning report on sexism in AI has been issued by the AI Now Institute at New York University: West, S.M., Whittaker, M. and Crawford, K. (2019). Discriminating Systems: Gender, Race and Power in AI. AI Now Institute. Retrieved from https://ainowinstitute.org/ discriminatingsystems.html. Does this suggest that diverse employees might design a better product? It does. Chinese facial recognition programs claim to be more skilled than the Western versions and are already meting out good citizen points for good behavior, demerits for undesirable behavior. We do not know if they're more skilled, or whether individual Chinese have protested that the system has been unfair to them. We do know that such social systems in the West would be deeply flawed by gender and racial bias. See too work showing how predictive algorithms in the justice system unfairly and inconsistently target defenants. Retrieved from https://news.harvard.edu.gazette/story/2018/05/grad-discovers-a…medium=email&utm_campaign=Daily Gazette 20180530

[3] The Googlers who began the walkout cheerfully used Google-invented tools to organize and recruit for it.

Silicon Valley's extreme sexism, racism, and ageism are a bitter disappointment to me. I long for that to change. I long for the kinds of projects that would do credit to us all and yes, slowly eradicate bias. Instead, right now we face unaccountable algorithms that "improve" themselves opaquely and top executives at a major firm like Facebook (an AI company if ever there was one) who practice and cover up grossly harmful social behavior by their engines.

On machine learning, writer and longtime software engineer Ellen Ullman warns, "In some ways we've lost agency. When programs pass into code and code passes into algorithms and then algorithms start to create new algorithms, it gets farther and farther from human agency. Software is released into a code universe which no one can fully understand" (Smith, 2018).

Yet alongside my naïve early hopes, my unexamined belief that pursuing more intelligence was like pursuing more virtue, my disappointments with the social aspects of Silicon Valley, I sensed very early that AI was momentous. I wanted to bear witness. To this day, for all my present deep misgivings, I'm thrilled and grateful that I could.

2.

In the mid-teens of the 21st century, a startling efflorescence appeared of declarations, books, articles, and reviews. (Typical titles: "The Robots Are Winning!" "Killer Robots are Next!" "AI Means Calling Up the Demons!" "Artificial Intelligence: Homo sapiens will be split into a handful of gods and the rest of us.") Even Henry Kissinger (2018) tottered out of the twilight of his life to declare that AI was the end of the Enlightenment, a declaration to give pause for many reasons.

The profound, imminent threat AI made to privileged white men caused this pyrexia. I laughed to friends, "These guys have always been the smartest one on the block. They really feel threatened by something that might be smarter." Because most of these priviledged white men admitted AI had done good things for them (and none of them so far as I know was willing to give up his smartphone), they brought to mind St. Augustine: "Make me chaste, oh Lord, but not yet."

Very few women took this up the same way (you'd think we don't worry our pretty heads). One who did, Louise Aronson, a specialist in geriatric medicine (2014), dared to suggest that robot caregivers for the elderly might be a positive thing, but Sherry Turkle (2014), another woman who responded to Aronson's opinon piece in *The New York Times* with a letter to the editor, worried that such caregivers only *simulated* caring about us. That opened some interesting questions about authentic caring and its simulation even among humans, but didn't address the issues around who would do this caregiving and how many of those caregivers society could afford.

As I read this flow of heated declarations about the evils of AI, ranging from the thoughtful to the personally revealing to the pitifully derivative—a Dionysian eruption if ever there was one—I remembered the brilliant concept, described and named by the film critic, Laura Mulvey, in 1975: *the male gaze*. She coined it to describe the dominant mode of filmmaking: a narrative inevitably told from a male point of view, with female characters as bearers, not makers, of meaning. Male filmmakers address male viewers, she argued, soothing their anxieties by keeping the females, so potent with threat, as passive and obedient objects of male desire. (The detailed psychoanalytic reasoning in her article you must read for yourself.)

In many sentences of Mulvey's essay, I could easily substitute AI for women: AI signifies something other than (white or Asian) male intelligence and must be confined to its place as a *bearer* not a *maker* of meaning. To the male gaze, AI is object; its possible emergence as autonomous subject, capable of agency, is frightening and must be prevented, because its autonomy threatens male omnipotence, male control (at least those males who fret in popular journals and make movies). Maybe that younger me who hoped AI might finally demolish universal assumptions of male intellectual superiority was on to something.

The much older me knows that if AI poses future problems (how could it not?) it already improves and enhances human intellectual efforts and has the potential to lift the burden of petty, meaningless, often backbreaking work from humankind. *Who does a disproportionate share of that petty, meaningless, backbreaking work?* Let a hundred Roombas bloom.[4]

[4] Journalist Sarah Todd wrote "Inside the surprisingly sexist world of artificial intelligence" (Quartz, October 25, 2015) about the sexism and lack of diversity in AI. The piece suggests women won't pursue AI because it de-emphasizes humanistic goals. Maybe public fears about the field are because of the homogeneity of the field, she went on. To close the gap, schools need to emphasize the humanistic applications of AI. And so on. Although many applications of AI grow out of a sexist culture and reflect that, readers of this history can also see the fallacies in Todd's argument. AI started out as a way of understanding human intelligence. That continues to be one of its major goals, which is why it partners with psychology and brain science. Its humanistic goals are central, whether to understand intelligence or to augment it. But all scientific and technological fields save, perhaps, the biological sciences, could use more women practitioners and more people of color. That is being addressed in many places and many ways, beyond the scope of this book, but one example is the national nonprofit AI4All, launched in 2017 by Stanford's Fei-Fei Li and funded by Melinda Gates, which aims to make AI researchers, hence AI research, more diverse. The 2019 report from NYU says this is not enough (West et al., 2019).

But the handwringing said that people were at last taking AI seriously.

3.

Another great change I've seen is the shift of science from the intellectual perimeters of my culture to its center. (Imagine C. P. Snow presenting his Two Cultures manifesto now. Laughable.) These days, not to know science at some genuine level is to forfeit your claims to the life of the mind. That shift hasn't displaced the importance of the humanities. As we saw with the digital humanities—sometimes tentative, sometimes ungainly, the modest start of something profound—the Two Cultures are reconciling, recognizing each other as two parts of a larger whole, which is what it means to be human. Not enough people yet know that a symbol-manipulating computer could be a welcome assistant to thinking, whether about theoretical physics or getting through the day.

AI isn't just for science and engineering, as in the beginning, but reshapes, enlarges, and eases many tasks. IBM's Watson, for instance, stands ready to help in dozens of ways, including artistic creativity: the program ("he" in the words of both his presenter and the audience) was a big hit at the 2015 Tribeca Film Festival when it was offered as eager colleague to filmmakers (Morais, 2015).

At the same time, AI also complicates many tasks. If an autonomous car requires millions of lines of code to operate, who can detect when a segment goes rogue? Mary Shaw, the Alan J. Perlis professor of computer science and a highly-honored software expert, worries that autonomous vehicles are moving too quickly from expert assistants beside the wheel and responsible for oversight, to ordinary human drivers responsible for oversight, to full automation without oversight. She argues that we lack enough experience to make this leap. Society would be better served by semi-autonomous systems that keep the vehicle in its lane, observe the speed limit, and stay parked when the driver is drunk. A woman pushing a bike, its handles draped with shopping bags, was killed by an autonomous vehicle because who anticipated *that*? If software engineering becomes too difficult for humans, and algorithms are instead written by other algorithms, then what? (Smith, 2018). Who gets warned when systems "learn" but that learning takes them to places that are harmful to humans? What programming team can anticipate every situation an autonomous car (or medical system, or trading system, or. . .) might encounter? "Machine learning is inscrutable," Harvard's James Mickens says (USENIX, 2018). What happens when you connect inscrutability to important real-life things, or even what he calls "the Internet of hate" also known as simply the Internet? What about AI mission creep?[5]

Columbia University's Jeanette Wing has given thought to these issues and offers an acronym: FATES. It stands for all the aspects that must be incorporated into AI, machine learning in particular: Fairness, Accountability, Transparency, Ethics, Security, and Safety. Those aspects should be part of every data scientist's training from day one, she says, and at all levels of activity: collection, analysis, and decision-making models. Big data is already transforming all fields, professions, and sectors of human activity, so everyone must adhere to FATES from the beginning.

But fairness? In real life, multiple definitions exist.

[5] The video in which Mickens' quote appears is mostly about the perils of machine learning, especially the hilariously sad story of Tay, Microsoft's chatbot, which had to be taken down from the Internet after 16 hours because of what it was learning from its training set, the gutter of the Internet.

Accountability? Who's responsible is an open question at present, but policy needs to be set, compliance must be monitored, and violations exposed, fixed, and if necessary, fined.

Transparency? Assurances of why the output can be trusted are vital, but we already don't fully understand how some of the technology works. That's an active area of research.

Ethics? Sometimes a problem has no "right" answer, even when the ambiguity might be encoded. Microsoft has the equivalent of an institutional review board (IRB) to oversee research (Google's first IRB fell apart publicly after a week), but firms aren't required to have such watchdogs, nor comply with them. According to Wing, a testing algorithm for deep learning, DeepXplore, recently found thousands of errors, some of them fatal, in fifteen state-of-the-art data neural networks in ImageNet and in software for self-driving cars. Issues around causality versus correlation have hardly begun to be explored.

Safety and security? Research in these areas is very active, but not yet definitive.

This could be important.

So I said again and again over my lifetime. Now we know. AI applications arrive steadily. Some believe we'll eventually have indefatigable, smart, and sensitive personal assistants to transform and enhance our work, our play, our lives. Researchers are acting on those beliefs to bring such personal assistants about: the Guardian Angel, Maslow, Watson. With such help, humans could move into an era of unprecedented abundance and leisure. Others cry halt! Jobs are ending! Firms and governments are spying on our every move! The machines will take over! They want our lunch! They lack human values! It will be awful![6]

Which will it be?

History says that however painful the transition, a major revolution—the agricultural, the industrial—has led, on balance, to greater advantages for humans in general. Historians know this, but it doesn't stop them from saying that this time may be different; this time we may be the losers. Because we may be. At the moment, we can only guess.[7] Kai-Fu Lee (2018), the eminent Chinese AI researcher and venture capitalist, reminds us that this revolution will be at least as important as the previous two—one plus one equals three, he puts it—and it's coming a lot faster. As AI gets better, the intellectual, ethical, and political issues it brings with it are starkly challenging, and we are imperfect vessels to wrestle with them.

[6] The cries of pain and alarm are too numerous to list. Privacy, meddling, reshaping our sense of ourselves as unique, and more. About the future job market, for example, books and articles abound. See, for example, the relatively optimistic book by Erik Brynjolfsson and Andrew McAfee, Race Against the Machine: How the Digital Revolution is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy (Ditigal Frontier Press, 2011) or the careful quantitative study from the University of Oxford by Carl Benedikt Frey and Michael A. Osborne, "The Future of Employment: How Susceptible are Jobs to Computerisation?" (September 17, 2013 and available via https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf). But later economists question these findings as mere extrapolation, with no allowance for new jobs that will be created. For example, Forbes.com's Parmy Olson wrote about a PwC report on AI in "AI won't kill the job market but keep it steady, PwC report says" (July 17, 2018).

[7] In 2017, Brynjolfsson and Tom Mitchell, an eminent AI researcher, chaired a National Academies panel that strongly recommended the development of new, more precise tools for measuring the impact of AI on jobs, including better data monitoring and analysis, the kinds of tools in common use with firms like Google and Facebook. Some take issue with AI research being sponsored by the Department of Defense, or the big, profit-making tech firms. These are arguable complaints, but the alternatives—to leave it to chance, the Chinese government, or some well-heeled foundation—don't seem more desirable. Dropping out is not an option.

For example, as I write, the U.S. government, at least, is hostile to compensating humans for the jobs they lose because of automation. All governments will face a totally novel reality in the future. Wise minds, human and machine, will need to consider their response to that reality, which might include income redistribution or might offer unimaginable new occupations, using every variety of intelligence—ethical, analytical, emotional, machine—that exists.

In 1969, the visionary Buckminster Fuller published *Operating Manual for Spaceship Earth*. He foresaw the rise of automation and the loss of jobs. He proposed a "life fellowship" in research and development, or in just plain thinking, for every human who became unemployed by automation. Of 100,000 such fellowships, perhaps only one might yield a breakthrough idea, but the idea would be so potent it would more than pay for the other 99,999. He imagined such fellowships—these days called universal basic income—would give everyone a chance to develop their most powerful mental and intuitive faculties. He imagined young people, frustrated by soulless jobs, might just want to go fishing. "Fishing provides an excellent opportunity to think clearly; to review one's life; to recall one's earlier frustrated and abandoned longings and curiosities. What we want everybody to do is to *think* clearly" (Fuller, 1978). From this he foresaw the advent of an age of great abundance and tranquility.

Our present landscape is mixed. The Chinese artist Cao Fei makes a film in a factory where the silence is eerie, humans nearly nonexistent. What should we feel? Sadness that, in the United States at least, decent wages for many have disappeared? Or gladness that hard, repetitive work is now the province of machines, that humans have been released from a grueling forty-hours per week of monotony? Both?

We know that if jobs are disappearing, the planet has no lack of tasks. A friend returns in 2018 from a trip to Guatemala, where, as a volunteer nurse, she helped in a dental clinic for the rural poor and unserved. Her watercolor images depict not only Guatemala's natural beauty, but also half a dozen of the rotted teeth that volunteer dentists had pulled from the mouths of young children, a small fraction of those pulled daily. The planet offers endless such tasks waiting to be done.

I spoke at the beginning of this book of the great intellectual structure underway, called computational rationality, which embraces intelligence wherever it's found: brains, minds, machines. I called it a new Hagia Sophia, temple of holy wisdom, because intelligence, as you've surely guessed, is one of the things I hold sacred.

But like everyone else, I cannot define nor measure intelligence. I think I know it when I see it (or its absence). Its definition is barely underway and its measurement a conundrum. This part is not yet science.

Late in his life, Harold Cohen said, "The whole of my history in relation to computing really has had to do with a change from the notion of the computer as an imitation human being to the recognition of the computer as an independent entity that has its own capacities which are fundamentally different from the ones we have." Many researchers share that view. Fundamentally different? Superficially different? We don't know. Perhaps computational rationality will clarify those questions, along with questions of what's appropriate for AI to take on, and what it mustn't.

We return to the male gaze. In a recent NYU study (West et al., 2019) a picture emerges of AI as a field in crisis around diversity across gender and race. Only 18% of authors at leading AI conferences are women, and more than 80% of AI professors are men. At Facebook, for example, women are only 15% of AI research staff; and only 10% at Google. (For people of color, the proportions are even worse.) The male gaze transmutes into the male stranglehold. This means that products—algorithms, heuristics—reinforce biases that follow historical patterns of discrimination. Face recognition programs have infamously failed to identify people of color, and the binary assumptions in assigning gender—a human subject is either male or female—are far too simplified and stereotypical to be effective across a variegated population. Efforts that emphasize training women, the report goes on, will probably benefit white women preponderantly (which is no reason why they shouldn't continue, but perhaps with different shapings).

The pushbacks against diversification are especially telling. One skeptic about diversity argues that "cognitive diversity" can be achieved by ten different white men in a room, so long as they weren't raised in the same household. Others argue that everyone is different from everyone else: isn't that sufficient diversity? The report calls this a "flattening" of diversity, making it into an empty signifier that ignores the lived experiences and documented history of women and minorities. Biological determinism ascends once more, not only in hiring practices, but in the systems that emerge from such a workforce.

To improve workplace diversity, the report makes eight recommendations, among them publishing compensation levels across all roles and job categories, broken down by race and gender; greater transparency in hiring practices; and incentives to encourage hiring and retention of under-represented groups. The report's introduction puts it in boldface: "The diversity problem is not just about women. It's about gender, race, and most fundamentally about power. It affects how AI companies work, what products get built, who they are designed to serve, and who benefits from their development." Let me add that the employees of such companies have been active in protest against these built-in bigotries, especially at Google. (But again, two women organizers of the Google protest claim to have suffered retaliation as a result.)

To address bias in AI systems, the report recommends four steps: transparency in systems and their uses; rigorous testing across the lifecycle of AI systems, including pre-release trials and independent monitoring; a multi-disciplinary approach to detect examples of bias; and thorough risk assessment before AI systems are designed.

It is all about power—power in the workplace, power in the products that emerge, power in society. No one with power gives it up without a fight.

A Dark Horse Comes Out of Nowhere

China and AI

1.

In 2017 the Chinese government officially announced its goal to achieve primacy as a center of AI innovation by 2030. That same year, a program called AlphaGo enjoyed a decisive victory over the best human Go player living, a nineteen-year-old Chinese man named Ke Jie. To Chinese people in Beijing's equivalent of Silicon Valley, this match was both an inspiration and a challenge, what Kai-Fu Lee, a leading AI researcher and venture capitalist, calls "China's Sputnik Moment."

In the winter of 2018, a number of Western news sources reported on this Chinese plan. For example, on February 8, 2018, the journal *Science*, in an article ominously titled "China's massive investment in artificial intelligence has an insidious downside," announced to English language readers the new Chinese initiative in detail. China was investing heavily in all aspects of information technology, from quantum computing to chip design. "AI stands on top of all these things," Raj Reddy was quoted as saying. He's right.

The article went on to praise the Chen brothers, whose AI chip, backed by the Institute of Computing Technology of the Chinese Academy of Sciences, has been the foundation of a new company, called Cambricon, already worth $1 billion dollars by 2017. China's State Council forecast that by 2030 the country's AI industry could be worth $150 billion. (We'd later learn that another special chip designed in China had allowed the Chinese to hack into major Western organizations.)

"China's advantages in AI go beyond government commitment," the *Science* article went on. Because of China's sheer size, vibrant online commerce and social networks, and scant privacy protections, the country is awash in data, the lifeblood of deep learning systems. Chen Yunji, the co-designer with his brother of a chip that can equal the performance of the 16,000 microprocessors Google Brain once needed to learn to identify a cat, also said that because AI is a young field, China benefits; AI's relative newness has encouraged "a burgeoning academic effort that has put China within striking distance of the United States" (Larson, 2018).

Chen's claim to a "burgeoning academic effort" in China is somewhat undermined by the reality that AI companies seek talent, offering salaries that no university can match. But the United States faces the same challenge, and some universities have met it by releasing American academics on semester- or year-long furloughs to make money and then welcoming them back to teach and do basic research. The Trump administration's anti-immigration policies have worsened the U.S. situation, sending promising students to Canada and Europe instead of the United States, which had traditionally welcomed and trained them (and often retained them).

China uses its AI prowess in ways that make Westerners deeply uncomfortable, although Western governments also employ such practices: surveillance (especially fine-grained facial recognition), censorship, battlefield decision-making, and autonomous weapons. This is part of the insidious downside the *Science* article refers to. But at least in the West, civil liberties organizations, such as the American Civil Liberties Union, protest by bringing suits against such U. S. government deployments of AI. Myriad organizations continuously examine AI applications and their ethics; and Europeans have gone even further to control social media with, for example, rules about the individual's right to be forgotten, or to have personal data deleted from public view. In Silicon Valley itself, valued employees are organizing to ask questions about the ethics of where their work might lead. Even a Chinese Internet conference in Wuzhen in 2018 featured sessions that grappled with unintended consequences of AI deployment: counterterrorism, data breaches, surveillance, issues around private enterprise (especially WeChat), and complicity with government intrusion (Zhong, 2018).

2.

Kai-Fu Lee is an outstanding AI expert who emigrated from Taiwan to the United States without a word of English when he was eleven, graduated with honors from Columbia University, and received his PhD at Carnegie Mellon under Raj Reddy. His doctoral dissertation was called Sphinx, the first large-vocabulary, speaker-independent continuous voice recognition program. It was so extraordinary that he had to publish his source code to convince others that the program was what it claimed to be. Lee has had years of experience in American firms such as Apple, Microsoft, and Google, and is now a venture capitalist based in Beijing, with investments in both China and the United States. His lifelong experience in both the East and West makes him an especially perceptive observer. [1]

In his 2018 book, *AI Superpowers: China, Silicon Valley, and the New World Order*, Lee describes the contrast between the victory of IBM's Deep Blue over the world's chess champion, Garry Kasparov, with the victory of a Google subsidiary, DeepMind's AlphaGo over the human Go champion, Ke Jie. Chess had been a brute force win, with specialized hardware and software applicable only to chess. The win was interesting, Lee says, but made little difference to the real world.

For the game of Go, brute force was useless. Go is so complex that to play it, let alone win, seems to require unique human intuition, human art. Thus when the program AlphaGo (later called AlphaZero) decisively won three out of four games against a nineteen-year-old human champion, Ke Jie, it had a real-world effect.

The machine's victory seized Chinese souls. Go was their own great game, one of the "four arts" all ancient Chinese scholars were expected to master, and the victory began what Lee calls an "AI frenzy" in China. To be sure, this was a moment of AI implementation rather than discovery, the frenzied apps built on earlier fundamental research done in the West. China's venture capitalists also responded to their government's challenge, and in 2017 were responsible for 48 percent of global AI venture capital funding, for the first time surpassing the United States.

[1] And disingenuous: when Kai-Fu Lee praises the face-recognition software the Chinese government employs, he tells Western audiences that if it were used at airports, terrorists could never get on planes. But Westerners are uncomfortably aware how authoritarian governments, East and West, can abuse such software.

It doesn't matter that the great AI research breakthroughs came from North America, the United States, and Canada, Lee says. Those fundamental breakthroughs (deep learning, for example) have offered China the chance to dream up apps that build on this research in unexpected ways. Yes, the Chinese can be accused of being copycats (even copykittens, as Lee jokes) but they know how to read their market, they're supple, they're local, their data is massive ("China is to data as Saudi Arabia has been to oil") and they make the Silicon Valley work ethic look slothful. Consider that in 2013, China had only two of the world's largest publicly traded tech companies, whereas the United States had nine. But by 2018, five years later, China had nine of the top twenty, and the United States eleven. Twenty years ago, China had none (Friedman, 2018).

Lee's tales of the Chinese entrepreneur wars make a gripping read—he compares them to the bloody gladiatorial combats in the coliseum, battles to the death, win—or die. Meanwhile, the Chinese government, seeing how vital AI is to the future Chinese economy, is "putting fingers on the scale"—that is, offering subsidies to venture capitalists and other promising enablers of AI apps.

Thus, without legacy systems (such as credit cards) to impede it, the Chinese Internet adapted quickly to mobile phones. People who couldn't afford desktops or laptops could easily acquire a cheap mobile phone, their introduction to life online.

The Chinese app WeChat, however, wanted to move beyond online and reach into people's offline lives. WeChat has thus become a super-app, a "remote control for life," that dominates not just online apps, but allows users "to pay at restaurants, hail taxis, unlock shared bikes, manage investments, book doctors' appointments, and have those doctors' prescriptions delivered to your door." Aside from contributing yet more data that AI algorithms can work on, it has blurred the distinction in China between online and real life (Lee, 2018).

Again, the Chinese government plays a big role. In its 2017 declaration aiming to achieve AI primacy by 2030, the central Chinese government laid out major economic goals for AI. (The Obama government, Lee points out, had issued a similar policy paper a few months earlier, but had the misfortune to release it the same week that presidential candidate Donald Trump's *Access Hollywood* tape came to public notice. Later, President Trump would propose to cut funds for AI research, but the Pentagon wasn't having any of that, and in fact Trump would eventually agree that AI research must be supported in the United States.

A more worrisome Chinese penetration is investment—the Chinese government is heavily invested in Silicon Valley ($35 billion over the last decade) buying startups to own their novel ideas. In 2018 the U.S. Congress passed legislation that expands government oversight on any foreign investments in "emerging technologies," and the power to block deals if they're considered unfavorable to domestic security (Canon, 2018).

3.

The central Chinese government set major national AI goals, but implementation details were left to provincial and municipal levels of government. Thus the original Avenue of the Entrepreneurs, near both Peking and Tsinghua Universities, became a model for cities all over China, helped along by grants and subsidies from the central government, followed by private capital, itself encouraged by government policy. Lee notes that in 2009, when he founded his venture capital firm, Sinovation, manufacturing and real estate dominated Chinese investing. By 2014, venture capital in AI quadrupled to $12 million, and then doubled again the following year (Lee, 2018).

China has also responded by beginning to construct an entire new city, Xiong'an, sixty miles south of Beijing, a "showcase city for technological progress and environmental sustainability," expected to reach a population of 2.5 million. It's built specifically to accommodate autonomous vehicles and environmental protection, with AI embedded in every nook and cranny. At a presentation at New York City's Asia Society in October 2018, Lee said that, after his book had gone to press, the city of Suzhou announced plans to rebuild a section of the ancient city, a two-level grid of streets where autonomous vehicles will be confined to the lower level, and human-driven vehicles, including bicycles, and pedestrians, will be on the top level.[2] Lee admits that other technologically themed cities in China haven't always succeeded, but many have, so how Xiong'an will fare is an open question.

The Chinese government's system of encouraging investments is intricate, but largely successful. Inefficient, scoff American investors. But effective, say the facts. "When the long-term upside is so monumental, overpaying in the short-term can be the right thing to do," Lee writes. "The Chinese government wanted to engineer

[2] I first visited Souzhou in 1981, where colorful boxes outside each pretty canal-side dwelling contained night soil to be collected as fertilizer for the surrounding countryside. Twenty years later, on another visit to a completely modernized Souzhou, I mentioned to my guide, a former mayor of the city and now a high-ranking national official, how much more picturesque I'd found the old city. He replied irritably: do you really think that was a superior way for ordinary people to live? No, I didn't. He was right to be irritated.

a fundamental shift in the Chinese economy, from manufacturing-led growth to innovation-led growth, and it wanted to do that in a hurry."

Lee's sharp-witted descriptions of the social changes AI has brought about in China are compelling. Since I'd lived through decades of the West's general sniffiness about computing in general and its resistance to AI in particular, I was astounded by the reactions of ordinary Chinese to this new science and its technology.

Almost overnight, great skepticism turned to avid fanaticism. Lee describes how difficult in the beginning it had been to recruit good minds for the startups funded by his venture capital firm, Sinovation Ventures, because a general view had long prevailed: one's children and one's spouse should aim for a lifetime job, an iron rice bowl, with the government. But once the government blessed AI startups, Lee found people knocking down Sinovation's door—literally, in one case—for the chance to work with him. "…Scrappy high school dropouts, brilliant graduates of top universities, former Facebook engineers, and more than a few people in questionable mental states."

The O2O—Online to Offline—Revolution in China was underway. It would turn online actions into offline services. E-commerce would make real-world services as convenient as things that arrived in boxes: hot food, a haircut, and a ride (the latter modeled on Uber, but done better for the Chinese, which drove Uber out of China and has become its rival in other countries). The list of services is awe-inspiring, and after the initial boom and bust (for the gladiatorial combat took place in O2O too), the urban service sector in China has been reshaped. WeChat, the super-app, offered a one-stop place to activate these services, in contrast to the constellation of apps that prevails in the United States.

Similarly, online services in China bundle related services, an approach Lee calls "going heavy." For example, whereas the U.S. model for apps "goes light," offering a single service such as handling restaurant orders (but leaving deliveries to the individual restaurants), the Chinese equivalent of Yelp not only rates restaurants but handles orders, delivers them, and is buying up gas stations and mo-ped repair shops. The Chinese equivalent of Airbnb lists homes, but also manages rental properties and handles the work of cleaning, stocking, and installing smart locks on each property. The long-run advantages of "going heavy" are in the data this yields about users' consumption patterns and personal habits. Mobile payments, of negligible cost to merchant and customer, turn a data edge into a commanding lead. Data is the fuel of machine learning, the present boisterous star of AI: the more data, the more usefully the algorithms can work.

4.

Advantage China? Maybe. Lee readily concedes that another breakthrough in AI, on the scale of deep learning, will change the game all over again, and it's likely that such a breakthrough will come from the freewheeling West, rather than the implementing East. But such breakthroughs usually occur only every few decades. (After deep learning was invented, nearly three decades passed before sufficient computing power arrived to make it useful.) Meanwhile, the myriad implementations based on past breakthroughs are led by China, improving on those apps by dogged trial and error, and informed by the vast data offered by the Chinese population's behavior. Not to mention old-fashioned spyware in a new-fangled form.

Although controversy surrounds the event, in October 2018, it was disclosed that a spy chip, barely the size of a grain of rice, somehow inserted by operatives in the People's Liberation Army, had been discovered on motherboards manufactured in China. The spy chip had evaded detection for some years and affected nearly thirty U.S. companies, including Apple, Amazon, a major bank, and undisclosed government agencies. Probably China's goal was access to high-value corporate and government secrets. For the record, Apple and Amazon called the story "erroneous," but experts guessed that the details passed the sniff test. The China-U.S. confrontation is much more than a friendly competition between commercial rivals, as Lee presents it (Robertson & Riley, 2018).

Yet Western observers have already begun to wonder aloud if the Chinese aren't on to something. Could the orthodox free-market ideology that has prevailed—indeed, achieved near cult status in the United States for decades—have its drawbacks? "I applaud the Chinese Government for supporting science and technology," Yasheng Huang, a professor of international management at MIT's Sloan School of Management says. "The U.S. should be doing that too" (Elstrom, Gao, & Pi, 2018). David Hoffman, the director of Intel's AI policy, talks about the development of an AI ecosystem. "One approach to that is, the market is just going to take care of that and develop that over time. Most other countries are saying, well, even if that is the case, we want to invest and to provide direction" (Jamrisko & Torres, 2018).

When I hear Silicon Valley libertarians bang on about how they want the government out of their businesses and their lives, I wonder at their ignorance of their own history. Without long, steady investment in both the Internet (which began as a military communications system) and in AI, there would be no Silicon Valley. The Defense Department was investing with discrimination in these technologies for much longer than any private investor would have tolerated with so little to show. The humans who made these investments on behalf of the Defense Department—on behalf of the American public—were visionaries and empiricists, not slaves of ideologies.

American reactions to China's ultra-fast rise in AI competence were predictable. It would be us vs. them, a clash of the AI Superpowers. Because some of this rivalry was arising just as the Trump administration was blaming China for everything including Original Sin, the rhetoric got heated. "Who will set the key rules of the global order in the 21st century?" Thomas J. Friedman asked, America, "the world's long-dominant economic and military superpower," or China, "its rising rival?" (2018).

"Nations are seeking to harness AI advances for surveillance and censorship, and for military purposes," wrote Christina Larson in *Science* (2018). Larson quoted several who fear this Chinese government investment in AI is less about delivering hot meals and haircuts and more about staying in power and stifling dissent. In *The San Francisco Chronicle,* war college instructor and retired U.S. Marine Thomas C. Linn (2018) writes:

> China is using artificial intelligence to build an Orwellian state. Smart cities track peoples' movements. China, netted with millions of cameras and facial and vehicle recognition systems, can rapidly identify individuals. Police wear facial recognition glasses that do the same. Biometric data provide even better identification. And people get social credit scores, which determine eligibility for loans, travel and more.

> This artificial-intelligence-enabled system enables political repression and strengthens autocratic rule.

All true. All distressing.

"China is reversing the commonly held vision of technology as a great democratizer, bringing more people freedom and connecting them to the world. In China, it has brought control" (Mozur, 2018). An experimental program in China even tracked the facial characteristics of tenth-grade students in a Hangzhou high school to detect their moods. "Educators in China have been sharply critical of the Hangzhou school, not only for invading students' privacy—neither they nor their parents were asked for consent—but for charging ahead with an unproven system that purports to improve student performance" (Lee, 2018). That program was suspended, at least temporarily.

5.

These are signs of a very different worldview from the Western ideal. "Today, few would confidently declare that the Chinese Communist party is on the wrong side of history," says Yuval Noah Harari, the Israeli historian and public intellectual, in his *21 Lessons for the 21st Century* (2018). Yet within 48 hours of the evening I heard Kai-Fu Lee paint a relatively benign picture of friendly if fierce commercial competition between the innovating West and the implementing East, factions in the Chinese government—"unharmonious voices"—were reported to be condemning the private enterprise that has brought China to such economic prominence, and supporting instead state-owned and -controlled enterprises, a return to old Marxist times now that prosperity has been achieved (Yuan, 2018). Those with long memories recall the Cultural Revolution and shiver.

A November 2018 report from Freedom House, the democratic watchdog organization, says the Chinese are exporting digital authoritarianism. Thirty-eight countries have installed large-scale telecommunications equipment from Chinese companies, allowing those countries to track citizens' everyday movements the way China tracks and controls its citizens, and furthermore allowing the Chinese to spy on the countries that have installed these systems. China even sponsors training for its international governmental customers in methods to control dissent and manipulate online opinions. The report cites the example of Uigars in western China, tracked by the Chinese government and sent to "re-education camps" (Abramowitz & Chertoff, 2018). But the Chinese and their government customers also must deal from time to time with the Dionysian, despite Appollonian rigidity. Joe and I were eye-witnesses and in personal danger during the decidedly Dionysian 1989 massacre in Tiananmen Square.

As China continues to develop systems that enable its authoritarianism, we could have three major internets, an outcome lamented in a lead editorial in *The New York Times* (Editorial Board, 2018). Eric Schmidt, Google's former chief executive, predicts that the global Internet will split into two within the next decade, a Web led by the United States and another by China, that one with fewer freedoms and greater censorship. Tim Berners-Lee, the inventor of the World Wide Web, thinks Europe should build its own Internet, protecting privacy and intellectual property in ways neither the United States nor China does.

6.

In *AI Superpowers*, Kai-Fu Lee is worried not about China versus the United States, but about global problems. He believes that the United

States has a great lead in innovations and China in applications and that the two will complement each other in AI for some time to come. But information and communications technologies differ from former disruptive technologies. The steam engine and electricity led to the loss of skills. The tasks of highly skilled master weavers, for example, were decomposed, and machines operated by much less skilled workers took their place. This change was hard on the master weavers (and transformed the life of the son of one of them, Andrew Carnegie) but raised a whole population of the unskilled into gainful employment.

With information and communications technologies (ICT), however, the results are more ambiguous, Lee observes. Worker productivity has steadily increased over the last thirty years, but those gains have not translated into wage or employment gains. Instead, we see increasing economic stratification; in the United States, the economic gains of ICT go to the top one percent of the population. ICT is often, though not always, biased in favor of high-skilled workers. “By breaking down the barriers to disseminating information, ICT empowers the world’s top knowledge workers and undercuts the economic role of many in the middle” (Lee, 2018).

This presents not technological but staggering social and political problems. Kai-Fu Lee and many others believe the AI revolution will be on the scale of the Industrial Revolution, but probably larger. We know it will be faster. AI will invade and enhance both muscle power and cognitive power, outperforming humans at many such tasks. But it will not ease the lot of the unskilled. It will take over tasks that, using data, can be optimized and tasks that don’t require human interaction. New jobs will be created, but probably not enough to make up for all lost jobs. Displaced workers can theoretically retrain for new jobs in fields that are difficult to automate, but this is highly disruptive and time-consuming (and training is so far largely in fields that are poorly paid).

Lee (2018) goes on to say that algorithms that perform white-collar work can be improved and disseminated quickly and cheaply, unlike the improvements that took place during the first two Industrial Revolutions in the 17th and 19th centuries and that were only fitfully adopted. He also argues that the presence of venture capital (VC) has changed the chancy patchwork of capital (private wealth, patronage, bank loans) that the first two revolutions relied upon. Instead VC numbers tell another story: global venture funding invested $148 billion in 2017, and AI start-ups accounted for $15.2 billion, a 141 percent increase over 2016. VCs will continue to seek every profit they can out of every appealing idea that AI researchers propose. AI is the first disruptive technology where China, a fifth of the world's population, equals the West, both in advancing and applying the technology. China's participation will accelerate AI.

Although Lee's book examines the effect of AI on jobs in persuasive detail, his biggest concern is the effects of the two AI superpowers, China and the United States, upon the rest of the world, driving an all but insuperable wedge, if AI is left unchecked, between the haves and the have-nots. AI is an inequality machine. Developing countries are losing the great, perhaps only advantage they've had: cheap labor. Put bluntly, China and the United States are going to divide up the world between them, even as the Pope once divided the world between Spain and Portugal, except this time it's for real.

Kai-Fu Lee's proposed solutions to AI's expected social impact are shaped by his own brush with mortality and deserve a reading in their own right. He proposes a fundamental rewriting of the social contract that rewards socially productive activities the same way the industrial economy rewarded economically productive activities. His specifics offer one concrete answer to my own vague longings that the AI bounty be fairly shared. Surely others will be imagined; if we're intelligent, carried out.

We face a new world, including a potential new conflict between two nation-state adversaries who wield power of colossal potency, a kind of power that has never before been seen or used on a global scale. This power could nullify past weapons of wars. The conflicts to come are economic and geopolitical, but also philosophical, and even spiritual.

Doing the Right Things

1.

Forty years ago, Herb Simon put the question to me: if human values could be perpetuated, in the form of "beasties" which carried on those values flawlessly, would I agree to that? Because I didn't say no at once, which surprised him, we never got to what we meant by human values. People still speak casually as if these values are immutable, when in fact they differ from culture to culture, and in any given culture, gradually (or sometimes suddenly) change.

Ethics evolve. Their conclusions are always provisional. We've enlarged and transformed human values, declared slavery an evil, condemned colonialism, and, within nation states, tentatively extended civil rights, educated our young, emancipated women. At times these values have moved in reverse. All this is incomplete.

We think we know: machines must *never*. . .and machines must *always*. . .

In all humility, we don't know. Ethical principles are proposed by professional associations, government consortiums and other groups, even, these days, by employees of AI-based firms, and must be tested, refined, amended, and tested again. Such principles might be transposed into law, but the law is a blunt, slow, and imperfect instrument for the subtle swiftness AI exhibits. As we've seen again and again, the great and sobering effect of AI upon intellectual projects is that it requires executable code. To work at all, an ethical AI requires, and will continue to require, bracing specificity. It will demand generation, testing, revisions, and exquisite sensitivity to outcomes.

The first widely reported test of ethics in AI has come with social media—should its masters have insouciantly allowed endless invasions into individual privacy merely to fatten their already staggering profits? Ignored manipulation of political propaganda by malicious actors (and then concealed that manipulation)? It's no secret that these malicious actors pose a real threat to electoral processes and democracy. Recent public rebellion against social media seems to say no, firms must exercise better control, but how are such firms to be forced to do so, even curbed? Much AI research is proprietary, in private hands. To make individual humans behave ethically is difficult enough. To demand that enterprises behave ethically raises different obstacles and questions—the profit motive seems to be as strong as the sex drive. And again, whose ethics?

A few years ago, I came down with a puzzling physical condition. After several doctors could name it but not suggest a cure, I considered turning my medical records over to a Silicon Valley firm that claimed to read millions, not hundreds, of papers of medical and scientific findings, and thereby discover a remedy. But I didn't turn those records over. Silicon Valley had already lost my trust.

Yet in 2018, a number of employees at Google risked their jobs—threatened to resign—by publicly protesting work pursued by the executives of the company, first against Department of Defense work, and then against an agreement for Google to get into the Chinese market by means of a product accommodating Chinese censorship (Conger & Metz, 2018). In these cases, Google executives reconsidered. One Defense contract, they decided, would not be renewed, and late in the year, they decided against competing at all for another $10 billion cloud computing contract at Defense. This was directly in response to employee concerns (Nix, 2018). Whether Google will enter the Chinese market, accommodating Chinese censorship rules, is unclear. Some 300 Microsoft employees protested their company's contracts with the federal Immigration and Customs Enforcement (ICE) agency, especially during the summer of 2018 when that agency was separating immigrant children from their parents as a presumed deterrent to illegal immigration.

Microsoft, however, has announced that the firm will sell to the Pentagon whatever advanced technology it needs for a strong national defense, and Amazon has joined in. China, as competitor, as adversary, as a political system that is inimical to democracy, hostile to human rights, looms ever in the calculations. At the same time, Microsoft's president, Brad Smith, has sued to protected customers' personal data from the government, and he's actively engaged in designing international agreements that would limit cyberweapons (Sanger, 2018).

I am of several minds about all this. After all these years, I feel attached to AI and share the revulsion young Google engineers and scientists feel about their work possibly used to harm other humans. I hopelessly wish this were not a world that forced us to confront such possibilities. Why not a different world, of cooperation and even kindness? But if you've read this far, you see that I was born in the pounding amidst a war and kept safe in my cot because women and men I would never know were willing to risk and often sacrifice their lives to defend me. I cannot be a pacifist. I love the country where I've lived and been an active citizen for most of my life. Do I wish it came closer to its ideals? Yes, deeply. But I want the opportunity, the time, to push harder at making my country match those ideals. For the foreseeable future, then, I want the best possible defenses of this imperfect land.

Plausible arguments exist that despite local conflicts, what has kept general world peace these last seventy years has been mutually assured destruction in case of nuclear war. If this is so—and how can we really know?—can the same principle of mutually assured destruction in cyberdefenses keep the future peace?

Kai-Fu Lee (2018), the scientist and venture capitalist whom we met in Chapter 30, contracted a grave illness that brought him close to death, and compelled him to think about many of these issues. His spiritual guide, Master Hsing Yun in Taipei, helped open his eyes. He concludes his book, *AI Superpowers: Silicon Valley and China*, by sharing what he learned and what he imagines for a righteous AI future. He'd like to see not the elimination of the professions by AI, but their transformation, for instance turning physicians into compassionate caregivers who no longer need to hold in their heads all the knowledge a physician must now hold, but who have easy access to universal health knowledge via AI. Such compassionate caregivers will be well trained, but they can be "drawn from a larger pool of workers than doctors and won't need to undergo the years of rote memorization that is required of doctors today. As a result, society will be able to cost-effectively support far more compassionate caregivers than there are doctors, and we would receive far more and better care". He envisions this transformation for many of the professions. "In the long run, resistance may be futile, but symbiosis will be rewarded." Unlike present human service jobs, these new jobs will be well paid, because the logic of private enterprise must alter. He argues that this alteration—a reversal of the emphasis on huge profits toward an emphasis on human service—is not just morally right, but self-protective. Yes, this kind of investment will need to accept linear rather than exponential returns, but such companies "will be a key pillar in building an AI economy that creates new jobs and fosters human connections."

Public policy must be involved too. Lee writes:

> I don't want to live in a society divided into technological castes, where the AI elite live in a cloistered world of almost unimaginable wealth, relying on minimal handouts to keep the unemployed masses sedate in their place. I want to create a system that provides for all members of society, but one that also uses the wealth generated by AI to build a society that is more compassionate, loving, and ultimately human.

Instead of a universal basic income, Lee proposes a social investment stipend, a decent government salary to those who invest their time and energy in activities that promote a kind, compassionate, and creative society. These would include three broad categories: care work, community service, and education, providing the basis for a new social contract that valued and rewarded socially beneficial activities in the same way we now reward economically productive activities. This could put "the economic bounty of AI to work in building a better society, rather than just numbing the pain of AI-induced job losses." He raises many questions about how such a change would be accomplished but believes the obstacles are not insurmountable.

Had Lee never experienced a terrifying diagnosis, harsh chemotherapy, the sharing of wisdom by his spiritual guide, and the love of his family, he writes, he might never have awakened to the centrality of love in the human experience. That made a simple universal basic income—a mere resource-allocation problem—seem hollow.

Meanwhile, I've already named a few problems AI presses us with: the loss of personal privacy, the lack of genuine diversity, the problem of making life better for everyone, not just the privileged few. Mark Zuckerberg, Facebook's founder and CEO, calls out for government regulation so it will be illegal for his firm and its rivals to gather, expose, or sell users' private information. Personal privacy, though tied up with AI, is clearly the forerunner of how we as a society will respond to AI in general. Now at only a stage of public uneasiness, policy is unformed. We need to work hard at this: policy around privacy is our tryout. We won't get it right the first time, nor the second. But each version will get closer to righteous balance.

Machines smarter than we are, which (or who) might not share our values—why would we do this? Why have we dreamed of doing it for so long? What happens to us when they arrive? How can we control them? Where should boundaries be drawn between human concerns and machine powers? Where *can* they be drawn?[1] The

[1] Oren Etizioni proposes three rules for AI "inspired by, yet develop further the 'three laws of robotics' that the writer Isaac Asimov introduced in 1942: A robot may not injure a human being, or through inaction, allow a human being to come to harm; a robot must obey the orders given it by human beings, except when such orders would conflict with the previous law; and a robot must protect its own existence as long as such protection does not conflict with the previous two laws." Etzioni's developed laws

cultural impulse toward them has been very long, very persistent: why? Might something be learned from such other intelligences?

Thinking fast, our first impulse, is no good here. AI brings on a host of the largest questions whose best answers will require collective thoroughness, collective experience, and time. Gradually, we might have to amend the social contract to include protection and responsibilities for humans and machines. A nonhuman intelligent entity raises wrenching questions that demand attention from well-trained ethicists, philosophers, spiritual leaders, computer scientists, historians, legal scholars, and many others, over a long period of time.

Can't we just pull the plug?

Who's we? Practically speaking, AI isn't conducted by some small group of scientists sequestered on a mountaintop in the New Mexico wilderness, to whom we can say imperiously: never mind. This is an international effort and has been for decades. If one nation (or group) decides to foreswear AI because of possible dangers, who else would cede the advantage? Any group that unilaterally decides to give up AI would find itself in shocking arrears, intellectually, socially, politically, and economically. As Ed Feigenbaum liked to say, AI is the manifest destiny of computing. To give up AI, a nation would have to give up computing altogether. Back to index cards in shoeboxes? Street corner public telephones (connected to a pathetic electro-mechanical system)? Seat-of-the-pants flying? The end of

are: An AI system must be subject to the full gamut of laws that apply to its human operator; an AI system must clearly disclose that it is not human; an AI system cannot retain or disclose confidential information without explicit pproval from the source of that information. Etzioni, "How to Regulate Artificial Intelligence," New York Times, September 1, 2017.

medical and biological research? Give up your smartphone? Ain't gonna happen.

Until now, only a handful of philosophers have taken AI seriously. Daniel Dennett, a philosopher at Tufts, always did—to the field's great benefit—and a few years ago, David Chalmers, a philosopher at NYU, also took up some of the questions AI raises.[2] Nick Bostrom, a philosopher and cognitive scientist at Oxford, has written *Superintelligence: Paths, Dangers, Strategies*,[3] a crackling study of when human-level or better AI might appear (probably by mid-century, he says, but it could be somewhat earlier or later); asking what that might mean for us, and suggesting strategies that could possibly shape AI to be desirable for humans, rather than a peril for them. He calls for "a bitter determination to be as competent as we can, much as if we were preparing for a difficult exam that will either realize our dreams or obliterate them." It is, he says, the essential task of our age.

One of them, certainly. Saving the planet looms at least as large. So does preserving democracy. So does relieving economic inequality.

I part ways with Bostrum on whether AI is unnatural or inhuman. Again, I think its creation is altogether natural, quintessentially human, inevitable. Complete control is a brittle illusion. But I agree with him and many others that a profound challenge looms. This challenge has never been a secret. In his 1976 address called "Fairy Tales," Allen Newell (1992) said long ago: there are trials to

[2] Or at least those alarmed by the notion of the Singularity, an idea that has always struck me as simplistic. David J. Chalmers, The Singularity: A Philosophical Analysis. Journal of Consciousness Studies. 17 (9-10) pp. 7-65, 2010. This was followed by responses from many: The Singularity: Ongoing Debate Part II, Journal of Consciousness Studies, 19, (7-8) 2012. My own contribution said in essence that people who actually faced the problem would be in the best position to deal with it.

[3] Bostrom's book also contains some hair-raising future scenarios of AI gone rogue.

overcome, dangers to brave, giants and witches to outwit. We must grow in virtue and in mature understanding; we must earn our prize. On the other hand, Eric Horvitz, an AI pioneer (an MD as well as a PhD) rightfully reminds us that without AI, 40,000 patients per year die right now from preventable medical accidents, and thousands more die daily on roads and in cars that ought to be safer, to name just a few problems that AI could prevent. In India, millions suffer from retinal pathologies leading to blindness, which AI could easily diagnose. The list goes on.

A report from Freedom House, a U.S. watchdog of worldwide democracy, exposes the repressive computer systems the Chinese are exporting and offers some remedies: sanctions upon companies that knowingly provide technology designed for repressive crackdowns, passage by U.S. legislators of the Global Online Freedom Act, "which would direct the secretary of state to designate Internet-freedom-restricting countries and prohibit export to those countries of any items that could be used to carry out censorship or repressive surveillance," along with requiring the companies that operate in repressive environments to release annual reports on what they are doing to protect human rights. Above all, the West needs to provide a better model of free information and protect citizens' data from misuse by governments, firms, and criminals (Abramowitz & Chertoff, 2018).

A group of European scientists has offered a sobering (if sometimes self-contradictory) survey of major things that might go wrong, encapsulated in the title: "Will democracy survive big data and artificial intelligence?" (Helbing et al., 2017). The group specifically rejects any top-down solution to our problems derived from AI or anywhere else. Social complexity continues to grow, they argue, and collective intelligence, formed from pluralism and diversity, has the best chance to solve unanticipated problems that arise. Not only is pluralism likelier to offer solutions to the problems such complexity brings, but like biological diversity, social diversity offers us resilience. Sadly, as we've seen, neither Silicon Valley nor its Chinese counterparts are diverse.

The Europeans also argue for informational self-determination and participation, improved transparency to achieve greater trust, reduced pollution and distortion of information, user-controlled information filters, digital assistants and coordination tools, collective intelligence, and the promotion of responsible citizen behavior in the digital world through digital literacy and enlightenment. The group details how such principles might be enacted.

2.

AI researchers themselves have formally been studying the social and ethical problems their field presents for more than a decade. In 2009, Eric Horvitz, Technical Fellow and Director at Microsoft Research Labs, was then president of the Association for the Advancement of Artificial Intelligence (AAAI). Believing it was time for AI to become proactive instead of reactive, he commissioned, and cochaired with Bart Selman, a Cornell professor of computer science, a meeting in Asilomar, California. Modeled on a 1975 gathering of molecular biologists who'd met to consider the long-term prospects and dangers of their research, the 2009 meeting was called the Presidential Panel on Long-Term AI Futures. Like the molecular biologists, AI researchers knew that they understood the possibilities and problems of AI better than nonspecialists, especially politicians, and wanted to assume responsibility for guidelines that would keep their research safe and beneficial.

This committee of AI professionals addressed popular beliefs that AI would produce disruptive outcomes, catastrophic or utopian. Panel experts were skeptical about these extremes—the Singularity, the "explosion of artificial minds,"—and assigned them a low likelihood but agreed that methods were needed to understand and verify the range of behaviors of complex systems, which would minimize unexpected outcomes. Efforts should be made to educate a rattled public and highlight the ways AI enhances the quality of human life.

In addition, the ways AI could help in the short term were proposed—by protecting individual privacy at the same time personal services are improved, improving joint human-AI tasks, and improving the ways machines explain their reasoning, goals, and uncertainties. AI might become more active in seeking out and preventing malicious uses of computing.

Ethical and legal issues in AI were scrutinized, especially as machines become more autonomous and active in what's called "high-stakes decisions," such as medical therapy or weaponry. Many are deeply offended by the idea of robots in warfare, for example, and want them banned absolutely, but for others, the issue isn't so clear-cut. We wish war would disappear and must work as hard as we can toward sustained peace. But if war doesn't disappear, will it be ethically more desirable to sacrifice human life on the battlefield, when robots might be suitable subsitutes? (I say nothing about the problems of understanding robot decision-making in real time, let alone controlling them.) And what about human relationships with systems that synthesize believable affect, feelings, and personality—those robots that read our faces and react "appropriately"? The panelists called for the participation of ethicists and legal scholars to help work through all these problems.[4] With some melancholy I note that a decade later, all these are still vexing problems (Horvitz, 2009).

Convinced that this was a vital effort and would be for years to come, in December 2014 Eric Horvitz and his wife Mary announced a large gift to Stanford University for a *hundred-year* study of the social and ethical issues surrounding AI, establishing AI100, a program to support ethicists, philosophers, AI researchers, historians, biologists, and anyone else whose work might be relevant. AI100 is overseen by a standing committee of distinguished AI researchers, its detailed agenda available on its website.[5] The study group's work was to be issued every five years. The Horvitzes believe that the social and ethical problems AI raises won't be solved once and for all—this must be a century-long effort, co-evolving with AI itself.

In late 2016, AI100 issued its first five-year study, *Artificial Intelligence and Life in 2030* (Stone et al., 2016). Among the topics taken up are technical trends and surprises, key opportunities for AI, the delays with technology transfer in AI, privacy and machine intelligence, democracy and freedom, law, ethics, economics, AI and warfare, the criminal uses of AI, AI and human cognition—the list is long.

The measured tone taken by the AI100 blue ribbon panelists, all of them eminent researchers in AI and related fields, didn't make headlines but might comfort the anxious. (Because much of the AI we encounter now in our everyday lives is based on research done twenty or more years ago by these scientists, they deserve our

[4] You're invited to lose sleep over an essay by Sarah A. Topol, "Attack of the Killer Robots," https://www.buzzfeed.com/sarahtopol/how-to-save-mankind-from-the-new-breed-of-killer-robots?utm_term-ronLOqqXlb#.hfJJRYY45D. Or a video that was making the rounds in 2017 from Stuart Russell: https://www.theguardian.com/science/2017/nov/13/ban-on-killer-robots-urgently-needed-say-scientists

[5] The AI100 website address is https://ai100.stanford.edu/

attention.) Yes, the science enables “a constellation of mainstream technologies that are having a substantial impact on everyday lives” such as video games, a bigger entertainment industry than Hollywood, practical speech understanding on our phones and in our homes and living rooms, and new power for Internet searches.

But—these technologies are highly tailored to specific tasks. General intelligence is very distant on the horizon. Thus the report focuses on specific domains: transportation, service robots, healthcare, education, low resource communities, public safety and security, employment and workplace, and entertainment. The report limits its scope to thirty years: the achievements of the last fifteen years and developments anticipated within the next fifteen years. The report makes some policy recommendations. It isn’t dry reading (in fact, it’s exciting); it just isn’t headline-making fantasies.

Some topics jump out: AI with neither commercial nor military applications has historically been underfunded, but targeted incentives and funding could help address the problems of poor communities—beginning efforts are promising. The State of Illinois and the city of Cincinnati use predictive models to identify pregnant women who might be at risk for lead poisoning or sites where code violations are likely. AI task-assignment scheduling and planning techniques have been used to redistribute excess food from restaurants to food banks, communities, and individuals. AI techniques could propagate health information to individuals in large populations who would otherwise be unreachable.

In the workforce, AI seems to be transforming certain tasks, *changing* jobs rather than *replacing* them, and it creates new jobs, too. Humans can see disappearing jobs but have a harder time imagining jobs yet to be created. The report says:

> Because AI systems perform work that previously required human labor, they have the effect of lowering the cost of many goods and services, effectively making everyone richer at least in the aggregate. But as exemplified in current political debates, job loss is more salient to people—especially those directly affected—than diffuse economic gains, and AI unfortunately is framed as a threat to jobs than a boon to living standards. (Stone et al., 2016)

In the longer term, AI may be thought of as a radically different mechanism for wealth creation in which everyone should be entitled to a portion of the world's AI-produced treasures. "Policies should be evaluated as to whether they foster democratic values and equitable sharing of AI's benefits, or concentrate power and benefits in the hands of a fortunate few."

Changes, yes: living standards raised, lives saved. But no imminent threat to humankind is seen, nor is likely to develop soon. The challenges to the economy and to society itself, however, will be broad. A long section of the report is devoted to public policy issues around AI and is frank: without intervention, AI could widen existing inequalities of opportunity if access to the technologies is unfairly distributed across society. "As a society, we are underinvesting resources in research on the societal implications of AI technologies. Private and public dollars should be directed toward interdisciplinary teams capable of analyzing AI from multiple angles." Pressure for more and tougher regulation might be inevitable, but regulations that stifle innovation would be equally counterproductive.

The real impediments and threats aren't scientific but political and commercial. Sadly, we can expect few politicians to read or act wisely upon such a report right now. As for the private sector, which engineers and managers at Facebook had read these recommendations before they allowed adversarial actors to mount a wholesale attack on American political discourse in the 2016 elections? How did those engineers and managers then square with their consciences the act of concealing their knowledge until forced by investigative journalists to admit the facts?

Perhaps more firms should establish institutional review boards, like those of universities and research hospitals, to examine proposed research and protect the humans who might be harmed by such research or business practices, but that seems a weak response to grave threats. In a capitalist society, it seems the only way to get firms to change their ways is for customers to vote with their feet. But if customers don't know what the choices are, how are they to vote?

Meanwhile, AAAI, the professional AI group, has established an annual conference on AI, Ethics, and Society, which calls for papers on such topics as building ethically reliable AI systems, moral machine decision-making, trust and explanation in AI systems, fairness and transparency in AI systems, AI for the social good, and other germane topics. In 2019, a young Chinese company called Squirrel AI Learning, which specializes in AI to enhance human education, announced the establishment of a million dollar annual prize for artificial intelligence that benefits humanity, to be administered by the Association for the Advancement of Artificial Intelligence. Although Squirrel AI Learning's own commercial focus is on education (the firm's teacher plus AI programs had won, among other prizes, the 2018 annual innovation award from Bloomberg/ BusinessWeek) the firm's president declared that the prize is for AI innovations across all sectors, its eye-catching largesse—at the level of the Nobel and the Turing—is meant to persuade the public that AI has great benefits, and to coax researchers to apply their intellectual resources toward such benefits.

"With great code comes great responsibility," begins an announcement of a new competition called the Responsible Computer Science Challenge (Mozilla blog, 2018), sponsored by a consortium that includes the Omidyar Network (eBay), Mozilla, Schmidt Futures (Eric and Wendy Schmidt, him of Google), and Craig Newmark Philanthropies (Craigslist). The challenge is to bring ethical education into computer science classes at the earliest moment and has two stages. The first seeks concepts from professors or teams of faculty and students for integrating ethics deeply into existing computer science courses. Stage 1 winners will receive $150,000 apiece to pilot their ideas. Stage 2 will support the dissemination of the best of Stage 1 programs, and winners will be announced in 2020 and receive $200,000 each to accomplish their goals. A distinguished committee of ethicists, computer scientists, and others will judge projects.

3.

Just after the establishment of AI100, a meeting was held in January 2015 in Puerto Rico, organized by the Future of Life Institute (FLI), a group of scientists and citizens concerned about issues presented by technology and especially AI. Elon Musk, the founder of Tesla Automobiles, put up $10 million to fund studies, although the Institute's agenda is less detailed than that of AI100. The public face of FLI roared extravagantly of threats, alarms, and catastrophes. Musk, Stephen Hawking, and Stuart Russell (an AI researcher at Berkeley and coauthor of the important textbook, *Artificial Intelligence: A*

Modern Approach, himself publicly comparing AI to atomic weapons) promoted an "open letter" in the summer of 2015, which argued for a ban on autonomous weapons, a term that became "killer robots" in the media. The letter soon collected tens of thousands of signatures and was presented to the U.N.[6] A later more detailed letter is on the Future of Life website: https://futureoflife.org/ai-open-letter.

Although some people think of these two efforts as competitive, the one slow and steady, the other a bright flash of male-gaze egos, Eric Horvitz believes in diversity. He helped organize the Puerto Rico program and describes each program as having related but distinct goals. The FLI program addresses fears of AI and safety, whereas the One Hundred Year Study "casts a broader focus on a wide variety of influences that AI might have on society. While concerns about runaway AI are a topic of interest at AI100, so is the psychology about smart machines," he told me.

[6] One response to that came from Evan Ackerman, editor in chief of IEEE Spectrum, in "We Should Not Ban 'Killer Robots' and Here's Why," which appeared in the July 29, 2015 issue: "The problem with this argument is that no letter, UN declaration, or even a formal ban ratified by multiple nations is going to prevent people from being able to build autonomous, weaponized robots," Ackerman began. He cited the toys already available at small cost and assumed market forces would only make them better and cheaper. Thus autonomous armed robots need to be made ethical, he argued, because we can't prevent them from existing. But they could make war safer: because they can be programmed, armed robots can perform better than humans in armed combat, and their accuracy and ethical behavior can be improved by reprogramming (which humans cannot be). "I'm not in favor of robots killing people. If this letter was about that, I'd totally sign it. But that's not what it's about; it's about the potential value of armed autonomous robots, and I believe that this is something that we need to have a reasoned discussion about rather than banning." Jerry Kaplan of Stanford wrote an op-ed piece in The New York Times, "Robot Weapons: What's the Harm?" (August 17, 2015), saying approximately the same thing. A later argument is that the United States has no choice but to increase its technological advantages, however fleeting and however difficult, given that so much AI research is for profit in the commercial sector. AI warfare is unprecedented, for which we are hardly prepared. Matthew Symonds, "The New Battlegrounds." The Economist, January 27, 2018.

The Future of Life Institute's original open letter was sloppily written with yawning loopholes had it somehow transmuted into law (enacted by whom? enforced by whom? with what sanctions? exceptions for national defense?). But psychologically, it represents something deeper. Those sensational earlier statements ("calling up the demons," "the end of the human race," "as dangerous as atomic weapons"), the open letter's language and promotion, are altogether Dionysian in the way Nietschze meant the term: that human embrace of the irrational, the extreme, the ecstatic and destructive, the terrifying darkness of the human psyche.[7] But Nietzsche was at pains to remind us that the Dionysian is as much a part of dualistic human nature as the Apollonian: rational, measured, illuminated, pursuing beauty and joy. Our truth is we are—and need—both. Six months after its establishment, the FLI reported distributing $7 million in grants to fund research toward its goals.

The Apollonian AI100 at Stanford will not be spending a million dollars a year funding incoming proposals. Horvitz told me:

> We have a different model. On funding levels, there have been offers of engagement and deeper funding to AI100 from others, including corporate sponsors. I've additionally suggested raising the level of funding by upping my own philanthropy. So far, the committee has come back with "stand by—we don't need additional funding," as they believe they have enough to work with—with the endowment (which

[7] For a deeper look at the founding and goals of the Future of Life Institute, along with some interesting scenarios of a future saturated with AI, see Max Tegmark's admirable Life 3.0: Being Human in the Age of Artificial Intelligence (2017). Recall Tegmark's schema of three stages of life: Life 1.0 is simple biological evolution, which can neither design its hardware or its software during its lifetime and can change only through evolution. Life 2.0 can redesign much of its software (humans can learn complex new skills, like languages or the professions, and can update their worldview and goals). Life 3.0, which doesn't yet exist on Earth, can dramatically redesign not only its software but its hardware as well instead of being delayed by evolution over generations.

> is actually set up to fund studies of a thousand years and beyond). My sense is that it's not the sheer funding level that's important, but the ideas, scholarship, programs, and balance, and the people attracted to and sought out by the project.

One more way to disinfect with sunshine: the eponymous Craig Newmark, of Craigslist, has given $20 million to a startup website called The Markup, whose purpose is to investigate technology and its effect on society (Bowles, 2018). Its editors are Julia Angwin, part of a Pulitzer-prize winning team at *The Wall Street Journal*, and data journalist Jeff Larson. They both also worked for ProPublica. (Three million dollars had also been raised from several other foundations, including the Ethics and Governance of Artificial Intelligence Initiative.) The site would employ programmers and data scientists as well as journalists with three initial focuses: how profiling software discriminates against the poor and other vulnerable groups; Internet health and infections, like bots, scams, and misinformation; and the awesome power of tech companies. Each editor was experienced in examining algorithms ("increasingly . . . used as shorthand for passing the buck," Larson said) for unintentional biases—showing, as they did, how criminal sentencing algorithms were unintentionally racist, how African-Americans are overcharged for auto insurance, and how Facebook allowed political ads that were actually scams and malwear.

This seemed a grand idea until a year later, Julia Angwin was forced to resign, and a majority of the newly-hired staff resigned with her in protest. The future of The Markup remains to be seen.

In *On Liberty*, in 1859, John Stuart Mill wrote: "It is as certain that many opinions, now general, will be rejected by future eyes, as it is that many once general, are rejected by the present." Thus the time scale of AI100 is significant.

These major efforts are complemented by independent queries in many countries—at universities in departments of philosophy, law, and computer science, and in think tanks like the United Kingdom's Center for the Study of Existential Risk at Cambridge University or the National Academies in the United States. Silicon Valley itself has put together the Open AI not-for-profit research company "that aims to carefully promote and develop friendly AI in such a way as to benefit humanity as a whole. The organization aims to freely collaborate with other institutions and researchers by making its patents and research open to the public. It's supported by over $1 billion in commitments, but that sum will be spent over a long period."[8] I've mentioned the annual Conference on AI, Ethics, and Society.

We humans *are* thinking about it. Whether we or our policymakers are up to the task of making it work for us in terms we value is another question. Whether engineers and managers think about it, carried away as they might be by the next big thing, isn't apparent, though we've seen lively protests erupt from people who work in Silicon Valley. The basic researchers in the field have been examining the social implications of field for a while and continue to do so.[9]

A consensus might possibly build to halt AI, but to repeat, that

[8] Melinda Gates, the philanthropist, and Fei-Fei Li, on leave as head of the Stanford University AI Lab, and acting as chief scientist of artificial intelligence and machine learning for Google Cloud, have recently formed AI4All, aimed to bring much more human diversity into AI research: more women, more people of color. "As an educator, as a woman, as a woman of color, as a mother, I'm increasingly worried. AI is about to make the biggest changes to humanity, and we're missing a whole generation of diverse technologists and leaders," says Li.

[9] In his 2017 presidential address, Thomas Dietterich, the president of AAAI, laid out a plan arguing that AI technology is not yet robust enough to support emerging applications in AI and proposed steps to remedy this. See "Steps Toward Robust Artificial Intelligence" in the Fall 2017 issue of AI Magazine.

consensus must be worldwide and deeply thought out over a long period. It can't be merely the snap judgment of the privileged. At the very least, it's unseemly for the privileged to bar research when AI might provide knowledge, abundance, and ease for that great majority on the planet who are now without. Self-righteousness is not sufficient: in the past, great ethical thinkers ferociously supported slavery, misogyny, racism, and homophobia, to name just a few of the ethical stances we've tried to evolve beyond. But that evolution took time and is incomplete.

A good ethical stance attends to both the inner and outer worlds. The outer world has the goal of seeing that the stance is effective: policymakers must not only be made aware of any consensus, but be persuaded to act on it. The inner world examines itself to be sure that its members are truly diverse and representative of the constituencies that will be affected. Whose assumptions about reality are included? Members of any panels or committees need to be deeply probed to guard against outcomes that are merely personal preferences writ large.

A good ethical stance also distinguishes among reality now, the normative, and the desirable. It considers what is, what ought to be (the content), and how to achieve what ought to be (strategies). Emergencies of the moment can suck up resources and contract the scope of the ethical stance instead of expanding it.[10]

When it comes time to turn decisions over to the experts, we need to know who they are. What are their goals? What's their individual character? Are they worthy of our trust? If ever a project presented

[10] I'm grateful for thought provoking and yes, entertaining, talk on this topic with Larry Rasmussen, the Reinhold Niebuhr Professor Emeritus of Social Ethics, Union Theological Seminary.

itself as perfect for a synthesis of the humanities and the sciences, this surely qualifies.

We're on a long, difficult, but exhilarating journey.

This Could Be Important

1.

But suppose AI's future is something else? Kevin Kelly, the founding editor of *Wired* magazine and a perceptive observer of technology for more than four decades, wrote in *Wired:*

> The AI on the horizon looks more like Amazon Web Services—cheap, reliable, industrial-grade digital smartness running behind everything, and almost invisible except when it blinks off. This common utility will serve you as much IQ as you want but no more than you need. Like all utilities, AI will be supremely boring, even as it transforms the Internet, the global economy, and civilization. It will enliven inert objects, much as electricity did more than a century ago. Everything we formerly electrified we will now cognitize. This new utilitarian AI will also augment us individually as people (deepening our memory, speeding our recognition) and collectively as a species. There is almost nothing we can think of that cannot be made new, different, or interesting by infusing it with some extra IQ. In fact, the business plans of the next 10,000 startups are easy to forecast: Take X and add AI. This is a big deal, and now it's here. (Kelly, 2014)

Five years after Kelly's predictions, this is about how AI seems.

Cheap parallel computation, big data, and better algorithms have brought us here, says Kelly. Google, for example, uses our daily searches to train its computers. The neural network model of computing suddenly has specialized chips (originally invented for games) that can do in a day what traditional processors needed several weeks to compute. Big data provides what's needed for computers to train themselves (although we've already seen the built-in problems with big data). Better algorithms have been developed over the last few decades to take perceptions from the lowest to the highest and most abstract levels of machine cognition—deep learning. But we must remember that present machine learning works only in a single domain, and only where an objective answer exists. It cannot cross domains; it cannot work at all if the initial conditions change even slightly.

Kelly goes on to envision such kinds of AIs as "nerdily autistic," dedicated exclusively to the single job at hand, whether that's driving a car or diagnosing and curing disease: focused, measurable, specific. "Nonhuman intelligence is not a bug, it's a feature." A new form of intelligence will think about manufacturing, food, science, finance, clothing, or anything else, differently. "The alienness of artificial intelligence will become more valuable to us than its speed or power." Kelly's observation recalls my journal entry of November 3, 1974: "I've come a long way from the time when I took offense at the idea of computers writing novels. Now I think I'd welcome a new form of intelligence to live in parallel with us."

Kelly's skepticism about a general-purpose machine intelligence is shared by William Regli, in 2017 the acting director of the Defense Advanced Research Projects Agency:

> The fact is, despite enormous individual engineering advances in recent years, we remain woefully inadequate when it comes to the art of

> design—the enigmatic and still largely unautomated process of synthesizing multiple elements into final products. (Regli, 2017)

In late 2018, Ed Finn, the founding director of the Center for Science and the Imagination at Arizona State University repeated—perhaps unwittingly—what John McCarthy once called "the literary problem," that our stories about future AI conform to literary conventions, contain heroes and villains, and the villain is nearly always AI, which prevents us from thinking seriously about a collaborative AI future, already here. Why a zero-sum competition? Finn asks. He wants to see holistic thinking about AI, bringing together science fiction writers, technologists, and policy makers.

2.

This book has been about humans, not machines. Humans were always my main interest. As it happens, AI's coming of age, if not yet its full maturity, has paralleled my own life. It gives me pause to think I've been acquainted with AI from the time it was a cozy fraternity of a few to now, when AI is in nearly every corner of our lives. So this book is not only a quest saga, but a coming of age story, of both a scientific field and of a naïve young woman, now slightly wiser, decidedly older. I was an undergraduate in the humanities, who bumped into AI early in its life and mine, had long conversations with its begetters, and warmed to their enthusiasm and optimism. I'd spend much of my life pulling on the sleeves of serious thinkers, trying to tell them that this—artificial intelligence—could be important.

I've offered a personal story here because, as I said at the outset, it's the particulars that illuminate: personalities, friendships, enmities, chance, context. To grasp these early times, abstractions wouldn't do.

The scientists who created AI, the scientists who push it forward, drew me to write about them, to stand as witness: they were and are brave, intellectually daring women and men, the early ones attacked and derided who, unfazed, went about changing the world. They deserve to be remembered as more than names of awards or carved on buildings.

You've seen that the future of AI is sometimes conceived as a wise Jeeves to our mentally negligible Bertie Wooster selves. *"Jeeves, you're a wonder." "Thank you sir, we do our best."* Watson, the Guardian Angel, Maslow, and its helpful bretheren want to be our car drivers, our financial and medical advisors, our teachers, our long-range planners, our colleagues—not our masters. This is an appealing picture, the human race riding effortlessly into the future in the slipstream of its own intelligent machines.

As one task after another falls to machines, we'll ask ourselves what human beings are, Kelly says. "The greatest benefit of the arrival of artificial intelligence is that AIs will help define humanity. We need AIs to tell us who we are." (2014)[1]

No, this is the continuing but newly refreshed task of the humanities, and it has already begun. As a teenager, I didn't ask who I was. I knew. I just didn't understand why the world didn't like or accept that. That's how I see any new definitions of us: accommodation to and illumination of our infinite variety.

[1] Kelly elaborates on these points in a later essay, "The AI cargo cult: The myth of a superhuman AI" (Retrieved from https://backchannel.com/the-myth-of-a-superhuman-ai-59282b686c62). Its main points are that intelligence is not a single dimension, and thus "smarter than humans" is meaningless, although dimensions of intelligence are not infinite. Humans do not have general-purpose minds and neither will AIs. Emulation of human thinking in other media (e.g., wetware) will be constrained by cost. Finally, intelligence is only one factor in progress.

We can't now say what living beside other, in some ways superior, intelligences will mean to us. Will it widen and raise our own individual and collective intelligence? In significant ways, it already has. Find solutions to problems we could never solve? Probably. Find solutions to problems we lack the wit even to propose? Maybe. Cause problems? Surely. AI has already shattered some of our fondest myths about ourselves and has shone unwelcome light on others. This will continue.

The future. It's been easy to resist writing breathless scenarios. Nothing ages faster nor makes the prophet seem so time-bound. As Jack Ma, the co-founder of the Chinese online service, AliBaba says, "There are no experts for the future. Only experts for yesterday."

When people ask me my greatest worry about AI, I say: what we aren't smart enough even to imagine.

3.

You might also recognize in all this ferment the two customary opposing views about AI—a catastrophe or a welcome blessing—an early theme from my own *Machines Who Think*: what I've called the Hebraic and the Hellenistic views of intelligence outside the human cranium. The Hebraic tradition is encoded in the Second Commandment: "You shall not make for yourself a graven image, or any likeness of any thing that is in heaven above, or that is on the earth beneath, or that is in the water under the earth."[2] We fear entertaining god-like aspirations, of calling down divine wrath for our overweening, illicit ambition. The Hellenistic view, on the contrary, welcomes (with cheer and optimism) outside help, the

[2] From Exodus 20:4, King James Version.

creations of our own hands—not that the dwellers in Olympus and their progeny didn't have problems.[3]

We already have a bitter taste of the dark side of AI. Russian bots and other software simulated human influencers and interfered with the U.S. national elections in 2016; our telecommunications and social media apps know our lives in granular, even embarrassing detail. The Chinese government, along with the Chinese army, runs deep learning algorithms over the search engine data collected about the users of Baidu, the Chinese equivalent of Google. Every Chinese citizen receives a Citizen Score, to determine whether they can get loans, jobs, or travel abroad (Helbing et al., 2017).[4] China is selling these systems to other countries. With all of us under surveillance, whether by our government or by firms, whether by manipulative individuals or scheming terrorists, how the economy and society are organized must change fundamentally. Kai-Fu Lee says we need to rewrite the social contract (2018). We do. Certainly we need to talk.

Let us talk too about the grand ideas in the Western tradition. What is thought? What is memory? What is self? What is beauty? What is love? What are ethics? Answers to these questions have up to now been assertions or hand-waving. With AI, the questions must be specified precisely, realized in executable computer code. Thus eternal questions are being examined and tested anew.

From the beginning, pioneering researchers in the field expected the machines would eventually be smarter than humans (whatever that meant), but they saw this as a great benefit. More intelligence was like

[3] The same division is evident in biological enhancement of human faculties. Some fear this very much; others think it would be a benefit. The combination of much smarter humans and much smarter machines is something to think about.

[4] Helbing, et al. should certainly be one of the texts we talk about.

more virtue. These early researchers were firmly in the Hellenistic tradition. They believed—and I do too, if you haven't guessed—that if we're lucky and diligent, we can create a civilization bright with the best of human qualities: enhanced intelligence, which is wisdom; with dignity, compassion, generosity, abundance for all, creativity, and joy, an opportunity for a great synthesis of the humanities and the sciences, by the people who specialize in each. Herb Simon liked to say that we aren't spectators of the future; we create it. A better culture, generously life-centered, ethically based yet accommodating infinite human variety, is a synthesizing project worthy of the best minds, human and machine.

We long to save ourselves as a species. For all the imaginary deities throughout history who've failed to save and protect us from nature, from each other, from ourselves, we're finally ready to substitute the work of our own enhanced, augmented minds. Some worry it will all end in catastrophe. "We are as gods," Stewart Brand famously said, "and might as well get good at it (1968)." We're trying. We could fail.

4.

Win or lose, we're impelled to pursue this altogether human quest. Some mysterious but profound yearning has led us here from the beginning. This is the deep truth of our legends, our myths, our stories. (It wants some explanation. This isn't exactly the joy of sex.) The search for AI parallels our innate wish to fly, to roam over and beneath the seas, to see beyond our natural eyesight. The quest takes us out of the commonplace, along a dark and perilous way, beset with tasks and trials, a collective hero's journey that all humans must undertake.

The tasks and trials we already see include the destruction of whole business models, the transformation of work (and thus for many, life's meaning), and faster-than-thought applications with unforeseen consequences. We face a possible, if unlikely, subjugation to the machines; a possible, if unlikely, destruction of the human race by AI. These seem to me remote, but trials we can't yet foresee will surely emerge. We hardly know how to meet the trials we can see. I quoted Herb Simon above: "We aren't spectators of the future; we create it." But often he also slightly misquoted Proverbs: "If the leaders have no vision, the people will perish."

For years I had these calligraphed words framed above my desk, a gift from my husband: "And wherefore was it glorious?"

I knew the rest of the passage by heart:

> Not because the way was smooth and placid as a southern sea, but because it was full of dangers and terror, because at every new incident your fortitude was to be called forth and your courage exhibited, because danger and death surrounded it, and these you were to brave and overcome. For this was it a glorious, for this was it an honourable undertaking. You were hereafter to be hailed as the benefactors of your species, your names adored as belonging to brave men who encountered death for honour and the benefit of mankind.

These are the words of the dying Dr. Victor Frankenstein, near the end of Mary Shelley's essential novel, *Frankenstein*. He cries out to a ship's crew that, during a hunt for the Northwest Passage, has been paralyzed with terror by the menacing ice. Yes, the words reflect ironically on his repudiation of his own creation of an extra-human intelligence. The deeper urgency, I believe, is his, and our, struggle to be brave, as we go where we must.

When the calligraphy and the rest of the passage it stood for was above my desk, I meant it for my own writing life, for my struggle to tell the world honestly, without exaggeration, about artificial intelligence. It can stand now for the human race's struggle to get the best from AI while curbing its dangers.

AI challenges and melds both art and science, and every other human resource. We've created something in our own image that might eventually surpass us, possibly destroy us as a species. With our grand, conspicuous, and shameful failures, maybe we deserve no better. But I'm still an optimist. Digital, yes; but *humanities*. We've never quite fallen out of love with ourselves, and it's been a great advantage. We might learn to collaborate with our smarter selves.

When I asked Marvin Minsky what his hopes were for AI, he replied: "That it step in where humans fail." Fair enough. I'd like AI, this once-in-human-history phenomenon, to enlarge our aspirations. The opportunity so far has often been squandered on relative trivialities, at least in the commercial sector. I long for us all to treat AI as the sacred trust it really is.

Wherefore was it glorious?

We've begun. Let us continue.

References

Abramowitz, M., and Chertoff, M. (2018, November 1). The global threat of China's digital authoritarianism. *The Washington Post.* Retrieved from https://www.washingtonpost.com/opinions/the-global-threat-of-chinas-digital-authoritarianism/2018/11/01/46d6d99c-dd40-11e8-b3f0-62607289efee_story.html

AI Index. (November 2017). *Artifical Intelligence Index: 2017 Annual Report.*Retrieved from https://aiindex.org/2017-report.pdf

Andersen, Richard. (2019, April). The intention machine. *Scientific American, 320*(4), pp. 24-31.

Aronson, L. (2014, July 19). The future of robot caregivers.*The New York Times.*Retrieved from https://www.nytimes.com/2014/07/20/opinion/sunday/the-future-of-robot-caregivers.html

Asia Society New York (Producer). (2018, October 1). AI Superpowers: A conversation with Kai-Fu Lee [Online video]. Retrieved from https://asiasociety.org/new-york/events/sold-out-ai-superpowers-conversation-kai-fu-lee

Berwick, R. C., and Chomsky, N. (2016). *Why Only Us: Language and Evolution.*Cambridge, MA: The MIT Press.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bowles, N. (2018, September 23). News site to investigate big tech, helped by Craigslist founder. *The New York Times.*Retrieved from https://www.nytimes.com/2018/09/23/business/media/the-markup-craig-newmark.html

Brand, S. (1968, Fall).*Whole Earth Catalog.*Retrieved from http://www.wholeearth.com/uploads/2/File/documents/sample-ebook.pdf

Brockman, J. (2014, November 14). The myth of AI: A conversation with Jaron Larnier. *Edge.*Retrieved from https://www.edge.org/conversation/jaron_lanier-the-myth-of-ai

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., and Schnapp, J. (2012). *Digital_Humanties.*Cambridge, MA: The MIT Press.

Canon, G. (2018, November 10). Why Silicon Valley is worried about U.S. plan to curb Chinese funds. *The Guardian.* Retrieved from https://www.theguardian.com/technology/2018/nov/10/silicon-valley-chinese-funding-trump-administration-pilot

Carey, B. (2019, April 24). Scientists create speech from brain signals. *The New York Times*. Retrieved from https://www.nytimes.com/2019/04/24/health/artificial-speech-brain-injury.html

Carley, K. M. (2015, October 24). Will social computers dream? *Welcome Back to the <Source> of It All.*Symposium celebrating the 50th anniversary of the Department of Computer Science at Carnegie Mellon University, Pittsburgh, PA.

Chalmers, D. J. (2010). The singularity: A philosophical analysis.*Journal of Consciousness Studies. 17 (9-10)* 7-65.

Cohn, G. (2018, October 25). AI art at Christie's sells for $432,500. *The New York Times.* Retrieved from https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html

Conger, K., and Metz, C. (2018, October 7). Tech workers now want to know: What are we building this for?*The New York Times*. Retrieved from https://www.nytimes.com/2018/10/07/technology/tech-workers-ask-censorship-surveillance.html

David, D. (2017). *Pamela Hansford Johnson: A Writing Life.*New York: Oxford University Press.

Davis, R., Shrobe, H., and Szolovits, P. (1993, Spring). What is knowledge representation? *AI Magazine, 14*(1), pp. 17-33. doi: https://doi.org/10.1609/aimag.v14i1.1029

Dennett, D. C. (2017). *From Bacteria to Bach and Back: The Evolution of Mind*. New York: W. W. Norton.

Dennett, D. C. (2013). *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton & Co.

Dowd, M. (1983, October 12). Columbia enters new era with computer center. *The New York Times.*Retrieved from https://www.nytimes.com/1983/10/12/nyregion/columbia-enters-new-era-with-computer-center.html

Dreyfus, H. (1972). *What Computers Can't Do: A Critique of Artificial Reason.*New York: Harper & Row.

Editorial Board. (2018, October 15). There may soon be three Internets. America's won't necessarily be the best. *The New York Times.*Retreived from https://www.nytimes.com/2018/10/15/opinion/internet-google-china-balkanization.html

Elstrom, P., Gao, Y., with Pi, X. (2018, July 10). China's technology sector takes on Silicon Valley. *Bloomberg News*. Retrieved from https://www.bloomberg.com/news/articles/2018-07-10/china-s-technology-sector-takes-on-silicon-valley

Eschner, K. (2018, February 14). Women were better represented in Victorian novels than modern ones. *Smithsonian.com.*Retrieved from https://www.smithsonianmag.com/arts-culture/what-big-data-can-tell-us-about-women-and-novels-180968153

Estorick, A. (2017, December 5).When the painter learned to program. *Flash Art*. Retrieved from https://flash—art.com/article/harold-cohen/

Fackler, M. (2017, November 19). Six years after Fukushima, robots finally find reactors' melted uranium fuel. *The New York Times.*Retrieved from https://www.nytimes.com/2017/11/19/science/japan-fukushima-nuclear-meltdown-fuel.html

Feigenbaum, E., and McCorduck, P. (1983). *The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World.*Reading, MA: Addison-Wesley.

Feigenbaum, E., and Shrobe, H. (1993, July). The Japanese national

Fifth Generation project: Introduction, survey, and evaluation. *Future Generation Computer Systems, 9*(2).

Finn, E. (2018, November 15). A smarter way to think about intelligent machines. *The New York Times*. Retrieved from https://www.nytimes.com/2018/11/15/opinion/killer-robots-ai-humans.html

Ford, K., Hayes, P., Glymour, C., and Allen, J. (2015, Winter). Cognitive orthoses: Toward human-centered AI. *AI Magazine, 36*(4), pp. 5-8. doi: https://doi.org/10.1609/aimag.v36i4.2629

Friedman, T. L. (2018, September 25). Trump to China: 'I own you.' Guess again. *The New York Times.* Retrieved from https://www.nytimes.com/2018/09/25/opinion/trump-china-trade-economy-tech.html

Fuller, R. B. (1978). *Operating Manual for Spaceship Earth*. New York: E. P. Dutton.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. (2015, July 17). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science, 349,*273-278. doi: 10.1126/science.aac6076

Gil, Y., Greaves, M., Hendler, J., and Hirsh, H. (2014, October 10). Amplify scientific discovery with artificial intelligence. *Science, 346*(6206), pp. 171-172. doi: 10.1126/science.1259439

Glasberg, E. (2014, February). Faculty Q&A: Dennis Tenen. *The Record, 39*(5), p. 7. Retrieved from https://archive.news.columbia.edu/files_columbianews/imce_shared/vol3905.pdf

Goldman, R. (2017, Febrary 14). Dubai plans a taxi that skips the driver and the roads. *The New York Times.*Retrieved from https://www.nytimes.com/2017/02/14/world/middleeast/dubai-passenger-drones.html

Guizzo, E., and Ackerman, E. (2015, June 9). How South Korea's DRC-HUBO robot won the DARPA robotics challenge. *IEEE Spectrum.*Retrieved from https://spectrum.ieee.org/automaton/robotics/humanoids/how-kaist-drc-hubo-won-darpa-robotics-challenge

Halberstam, D. (2012). *The Fifities.*New York: Open Road Media.

Harari, Y. N. (2015). *Sapiens: A Brief History of Humankind.*New York: HarperCollins.

Harari, Y. N. (2018). *21 Lessons for the 21st Century*. New York: Spiegel and Grau.

Helbing, D., Frey, D. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., . . . Zwitter, A. (2017, February 25). Will democracy survive big data and artificial intelligence? *Scientific American.* Retrieved from https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/

Hern, A. (2019, February 14). New AI fake text generator may be too dangerous to release, say creators. *The Guardian.*Retrieved from https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction

High, R. (2013, November 4). Teaching IBM's Watson how to think like a human. *Forbes.*Retrieved from https://www.forbes.com/sites/ibm/2013/11/04/teaching-ibms-watson-how-to-think-like-a-human

Hirschberg, J., and Manning, C. D. (2015, July 17). Advances in natural language processing. *Science, 349*(6245), pp. 261-266. doi: 10.1126/science.aaa8685

Hofstadter, D., and Dennett, D. C. (1981). *The Mind's I.*Basic Books: New York.

Holmes, D., and Winston, P. H. (2018, December 15). The genesis enterprise: Taking artificial intelligence to another level via a computational account of human story understanding [technical report]. *Computational Models of Human Intelligence.*Retrieved from http://dspace.mit.edu/handle/1721.1/119651

Hong, J. (2015, October).*Intelligent Agents for Helping Humanity Reach Its Full Potential.*Retrieved from CHIMPS Lab http://www.cmuchimps.org/publications/intelligent_agents_for_helping_humanity_reach_its_full_potential_2015

Horvitz, E. (2009). Asilomar Study on Long-Term AI Futures: Highlights of 2008-2009 AAAI Study: Presidential Panel on Long-Term AI Futures. Retrieved from http://www.aaai.org/Organization/presidential-panel.php

Humanties cross the digital divide. (2014, February). *The Record, 39*(5), p. 1. Retrieved from https://archive.news.columbia.edu/files_columbianews/imce_shared/vol3905.pdf

Iacocca, L., and Novak, W. (1984). *Iacocca: An Autobiography.*New York: Bantam Books.

James, H. (1909).*The Ambassadors.* New York: Charles Scribner.

Jamrisko, M., and Torres, C. (2018, June 6). America may need to adopt China's weapons to win the tech war. *Bloomberg Quint.* Retrieved from https://www.bloombergquint.com/technology/america-may-need-to-adopt-china-s-weapons-to-win-the-tech-war

Kahneman, D. (2011). *Thinking, Fast and Slow.*New York: Farrar, Straus, and Giroux.

Kamvar, S. (2012, September 15–2013, April 1).*Boundaries* [art exhibit]. Skissernas Museum | Museum of Public Art. Lund, Sweden.

Katsma, H. (2014, September). Loudness in the novel. *Pamphlets of the Stanford Literary Lab.* Retrieved from https://litlab.stanford.edu/LiteraryLabPamphlet7.pdf

Kelly, K. (2014, October 27). The three breakthroughs that have finally unleashed AI on the world. *Wired.*Retrieved from https://www.wired.com/2014/10/future-of-artificial-intelligence/

Kissinger, H. A. (June 2018). How the enlightenment ends. *The Atlantic*. Retrieved from https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/

Kolodner, J. L. (2015, Winter). Cognitive prosthetics for fostering learning: A view from the learning sciences.*AI Magazine, 36*(4), pp. 34-50. doi: https://doi.org/10.1609/aimag.v36i4.2615

Krakovsky, M. (2014, September). Q&A: Finding themes. *Communications of the ACM,* 57(9), pp. 104–105. doi: 10.1145/2641223

Lanier, J. (2013). *Who Owns the Future?*New York: Simon and Schuster.

Lanier, J. (2017). *Dawn of the New Everything*. New York: Henry Holt and Company.

Larson, C. (2018, February 8). China's massive investment in artificial intelligence has an insidious downside. *Science.*doi:10.1126/science.aat2458

Leavis, F. R. (2013). *Two Cultures? The Significance of C. P. Snow with Introduction by Stefan Collini.*New York: Cambridge University Press.

Lee, D. (2018, June 30). At this Chinese school, Big Brother was watching students—and charting every smile or frown. *Los Angeles Times*. Retrieved from https://www.latimes.com/world/la-fg-china-face-surveillance-2018-story.html

Lee, K . (2018). *AI Superpowers: China, Silicon Valley, and the New World Order*. Boston: Houghton Mifflin Harcourt.

Linn, T. C. (2018, August 28). Race to develop artificial intelligence is one between Chinese authoritarianism and U.S. democracy. *San Francisco Chronicle.*Retrieved from https://www.sfchronicle.com/opinion/openforum/article/Race-to-develop-artificial-intelligence-is-one-13189380.php

Lohr, S. (2018, October 15). M.I.T. plans college for artificial intelligence, backed by $1 billion. *TheNew York Times*. Retrived from https://www.nytimes.com/2018/10/15/technology/mit-college-artificial-intelligence.html

McCorduck, P. (1976, February 9). An introduction to the humanities with Prof. Ptolemy. *The Chronicle of Higher Education*. Reprinted in *How to Read Slowly,* James W. Sire, ed., Downers Grove, Ill.: InterVarsity Press, 1978; in *The Joy of Reading,* James W. Sire, ed., Portland, Oregon: Multnomah Press, 1984; and in the 2nd edition of Sire, *How to Read Slowly,* Wheaton, Ill.: The Harold Shaw Press, 1989.

McCorduck, P. (1990). *Aaron's Code*. New York: W. H. Freeman.

McCorduck, P. (2004). *Machines Who Think, 2nd Edition.*Natick, Massachusetts: A. K. Peters Ltd.

Mill, J. S. (2011, January 10). *The Project Gutenberg EBook of On Liberty.*Retrieved from https://www.gutenberg.org/files/34901/34901-h/34901-h.htm

Miller, A. I. (2014). *Colliding Worlds: How Cutting-Edge Science Is Redefining Contemporary Art*. New York: W. W. Norton.

Minsky, M. (1988). *The Society of Mind.*New York: Simon & Schuster.

Minsky, M. (2006). *The Emotion Machine.*New York: Simon & Schuster.

Morais, B. (2015, April 24). Watson's star turn at Tribeca. *The New Yorker.*Retrieved from https://www.newyorker.com/business/currency/watsons-star-turn-at-tribeca

Moretti, F., and Pestre, D. (2015, March). Bankspeak: The language of World Bank reports, 1946–2012. *Pamphlets of the Stanford Literary Lab.* Retrieved from https://litlab.stanford.edu/LiteraryLabPamphlet9.pdf

Mozilla blog. (2018, October 10). Announcing a competition for ethics in computer science, with up to $3.5 million in prizes [Blog post]. Retrieved from https://blog.mozilla.org/blog/2018/10/10/announcing-a-competition-for-ethics-in-computer-science-with-up-to-3-5-million-in-prizes/

Mozur, P. (2018, July 8). Inside China's dystopian dreams: A.I., shame and lots of cameras. *The New York Times.*Retrieved from https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html

Mulvey, L. (1975, Autumn). Visual pleasure and narrative cinema. *Screen,* 16(3), pp. 6-18.

Muskus, J. (2018, May 17). AI made these paintings. *Bloomberg Businessweek.*Retrieved from https://www.bloomberg.com/news/articles/2018-05-17/ai-made-incredible-paintings-in-about-two-weeks

Naughton, J. (2018, August 5). Magicial thinking about machine learning won't bring the reality of AI any closer. *The Guardian.*Retrieved from https://www.theguardian.com/commentisfree/2018/aug/05/magical-thinking-about-machine-learning-will-not-bring-artificial-intelligence-any-closer

Newell, A., and Simon, H. A. (1972). *Human Problem Solving*. Upper Saddle River, NJ: Prentice Hall.

Newell, A. (1981). The knowledge level. *Artificial Intelligence, 2*(2), pp. 1-33. doi: https://doi.org/10.1609/aimag.v2i2.99

Newell, A. (1990). *Unified Theories of Cognition: The William James Lectures, 1987.*Cambridge, MA: Harvard University Press.

Newell, A. (1992). Fairy Tales. *Artifical Intellence, 13*(2), pp. 46-48. doi: https://doi.org/10.1609/aimag.v13i4.1020

Nichols, P. (Director). (1991, May 10). The Computer Bowl III, Part 2 [Television series episode]. Janice del Sesto & Stewart Chiefet (Executive Producers), *Computer Chronicles.*San Mateo, CA: KCSM-TV. Retrieved from https://archive.org/details/episode_851

Nichols, P. (Director). (1991, May 3). The Computer Bowl III, Part 1 [Television series episode]. Janice del Sesto & Stewart Chiefet (Executive Producers), *Computer Chronicles.*San Mateo, CA: KCSM-TV. Retrieved from https://archive.org/details/computerbowl

Nilsson, N. (2010). *The Quest for Artificial Intelligence.* New York: Cambridge University Press.

Nix, N. (2018, October 8). Google drops out of Pentagon's $10 billion cloud competition. *Bloomberg News.* Retrieved from https://www.bloomberg.com/news/articles/2018-10-08/google-drops-out-of-pentagon-s-10-billion-cloud-competition

Noë, A. (2015). *Strange Tools: Art and Human Nature.* New York: Hill and Wang.

Overbye, D. (2014, October 3). Martin Perl, 87, dies; Nobel Laureate discovered subatomic particle. *The New York Times.* Retrieved from https://www.nytimes.com/2014/10/04/science/martin-perl-physicist-who-discovered-electrons-long-lost-brother-dies-at-87.html

Penrose, R. (1989). *The Emperor's New Mind.* New York: Oxford Universtiy Press.

Raghavan, P. (Interviewer). (2011, June 6). *Joe (Joseph) Traub oral history*[Transcription of video]. Oral History Collection (Catalog No. 102745087, Lot No. X6067.2011). Computer History Museum, MountainView, CA.

Regli, W. (2017, Fall). Design and intelligent machines. *AI Magazine, 38*(3), pp. 63-65. doi: https://doi.org/10.1609/aimag.v38i3.2752

Robertson, J., and Riley, M. (2018, October 4). The big hack. *Bloomberg Businessweek.* Retrieved from https://www.bloomberg.com/news/features/2018-10-04/the-big-hack-how-china-used-a-tiny-chip-to-infiltrate-america-s-top-companies

Rose, F. (2018, September 14). Frank Gehry's Disney Hall is technodreaming. *The New York Times.* Retrieved from https://www.nytimes.com/2018/09/14/arts/design/refik-anadol-la-philharmonic-disney-hall.html

Russell, S., and Norvig, P. (2010). *Artificial Intelligence: A Modern Approach, 3rd Edition.*New York: Pearson.

Sancton, J. (2014, April 24). The culture conversation. *Departures.*Retrieved from https://www.departures.com/art-culture/culture-watch/culture-conversation

Sanger, D. E. (2018, October 26). Microsoft says it will sell Pentagon

Artificial intelligence and other advanced technology. *The New York Times.* Retrieved from https://www.nytimes.com/2018/10/26/us/politics/ai-microsoft-pentagon.html

Schuessler, J. (2017, October 30). Reading by the numbers: When big data meets literature. *The New York Times.*Retrieved from https://www.nytimes.com/2017/10/30/arts/franco-moretti-stanford-literary-lab-big-data.html

Shapiro, G. (2014, February). Columbia people: Alex Gil. *The Record, 39*(5), p. 4. Retrieved from https://archive.news.columbia.edu/files_columbianews/imce_shared/vol3905.pdf

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129-138. doi: http://dx.doi.org/10.1037/h0042769

Simon, H. A. (1991). *Models of My Life.*New York: Basic Books.

Simon, H. A. (1996). *The Sciences of the Artifical, 3rd Edition.*Cambridge, MA: The MIT Press.

Simon, H. A. (1997). Allen Newell. *Biographical Memoirs*(pp. 141-172). Washington, D.C.: National Academy Press.

Smith, A. (2018, August 30). Franken-algorithms: the deadly consequences of unpredictable code. *The Guardian.*Retrieved from https://www.theguardian.com/technology/2018/aug/29/coding-algorithms-frankenalgos-program-danger

Solman, P. (Business and Economics Correspondent). (2012, July 10). As Humans and Computers Merge . . . Immortality? [News segment]. Winslow, Linda (Executive Producer), *PBS NewsHour.*Arlington, VA: WETA Public Broadcasting.

Solon, O. (2018, June 18). Man 1, machine 1: Landmark debate between AI and humans ends in draw. *The Guardian.*Retrieved from https://www.theguardian.com/technology/2018/jun/18/artificial-intelligence-ibm-debate-project-debater

Somers, J. (2018, December 28). How the artificial-intelligence program AlphaZero mastered its games. *The New Yorker*. Retrieved from https://www.newyorker.com/science/elements/how-the-artificial-intelligence-program-alphazero-mastered-its-games

Somers, J. (2017, September 29). Is AI riding a one-trick pony? *MIT Technology Review*. Retrieved from https://www.technologyreview.com/s/608911/is-ai-riding-a-one-trick-pony/

Stockton, F. R. (1895). The Lady or the Tiger? *A Chosen Few Short Stories* (pp.117-128). New York: Charles Scribner's Sons. Retrieved from https://www.gutenberg.org/files/25549/25549-h/25549-h.htm#tiger

Stoica, I., Song, D., Raluca, A. P., Patterson, D. A., Mahoney, M. W., Katz, R. H., . . . Abbeel, P. (2017, October 16). *A Berkeley view of systems challenges for AI.*(Technical Report No. UCB/EECS-2017-159). Retrieved from the Berkeley Electrical Engineering and Computer Sciences website: http://www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017-159.html

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., . . . Teller, A. (2016, September). *Artificial Intelligence and Life in 2030*(Report of the 2015-2016 Study Panel). Stanford University, One Hundred Year Study on Artificial Intelligence. Retrieved from https://ai100.stanford.edu/2016-report

Stone, R., and Lavine, M. (2014, October 10). The social life of robots. *Science 346*(6246). doi: 10.1126/science.346.6206.178

Strickland, E. (2014, February 28). Dismantling Fukushima: The world's toughest demolition project. *IEEE Spectrum.*Retrieved from https://spectrum.ieee.org/energy/nuclear/dismantling-fukushima-the-worlds-toughest-demolition-project

Strogatz, S. (2018, December 26). One giant step for a chess-playing machine. *The New York Times.*Retrieved from: www.nytimes.com/2018/12/26/science/chess-artificial-intelligence

Stuart, K. (2018, September 24). From superheroes to soap operas: Five ways video game stories are changing forever. *The Guardian.*Retrieved from https://www.theguardian.com/games/2018/sep/24/from-superheroes-to-soap-operas-five-ways-video-game-stories-are-changing-forever

Tattersall, I. (2014, September). If I had a hammer. *Scientific American, 311*(3), pp. 54–59.

Tegmark, M. (2017). *Life 3.0: Being Human in the Age of Artificial Intelligence.*New York: Alfred A. Knopf.

Terras, M., Nyhan, J., and Vanhoutte, E. (Eds). (2013). *Defining DigitalHumanities: A Reader*. London: Ashgate.

The evolving university. (2014, Spring). *Columbia Magazine,*pp. 28-31.

The tech giant everyone is watching. (2018, June 30). *The Economist, 427*(9098), p. 11.

Togyer, J. (2014, Summer). Institutional memories: Reflections on a quarter-century and more. *The Link, 8*(1), 17-26. Retrieved from https://www.cs.cmu.edu/sites/default/files/14-399_The_Link_Newsletter-May.pdf

Turkle, S. (2014, July 25). Letter: How . . . are . . . you . . . feeling . . . today? *The New York Times.*Retrieved from https://www.nytimes.com/2014/07/26/opinion/when-a-robot-is-a-caregiver.html

USENIX (Producer). (2018, August 16). USENIX Security '18-Q: Why do keynote speakers keep suggesting that improving security is possible? (YouTube video). Retrieved from https://youtu.be/ajGX7odA87k

Valiant, L. (2014). *Probably Approximately Correct: Nature's Algorithms for Learningand Prospering in a Complex World.*New York: Basic Books.

Wakabayashi, D., Griffith, E., Tsang, A., and Conger, K. (2018, November 1). Google walkout: Employees stage protest over handling of sexual harassment. *The New York Times*. Retrieved from https://www.nytimes.com/2018/11/01/technology/google-walkout-sexual-harassment.html

Wakabayashi, D., and Brenner, K. (2018, October 25). How Google protected Andy Rubin, “the father of the Android.” *The New York Times*. Retrieved from https://www.nytimes.com/2018/10/25/technology/google-sexual-harassment-andy-rubin.html

Wapner, Jessica (2019) “The Engineer using A.I. to Read Your Feelings,” March 29, 2019. Retrieved from https://onezero.medium.com/the-engineer-using-a-i-to-read-your-feelings-bc284343f02)

Weizenbaum, J. (1976). *Computer Power and Human Reason.* New York: W. H. Freeman.

Weizenbaum, J. (1983, October 27). The computer in your future. *The New York Review of Books.*Retrieved from https://www.nybooks.com/articles/1983/10/27/the-computer-in-your-future/

West, S.M., Whittaker, M. and Crawford, K. (2019). Discriminating Systems: Gender, Race and Power in AI. AI Now Institute. Retrieved from https//ainowinsitute.org/discriminatingsystems.html

Wilson, E. O. (2014). *The Meaning of Human Existence*. New York: Liveright Publishing Corporation.

Wing, J. M. (2006, March). Computational thinking. *Communications of the ACM, (49)*3, pp. 33-35.

Winston, P. H. (2011). The strong story hypothesis and the directed perception hypothesis. In AAAI Fall Symposium Series. Retrieved from https://www.aaai.org/ocs/index.php/FSS/FSS11/paper/view/4125

Wortham, J. (2014, July 26). When digital art is suitable for framing.

The New York Times Bits Blog.Retrieved from https://bits.blogs.nytimes.com/2014/07/26/when-digital-art-is-suitable-for-framing/

Yuan, L. (2018, October 3). Private businesses built modern China. Now the government is pushing back. *TheNew York Times*. Retrieved from https://www.nytimes.com/2018/10/03/business/china-economy-private-enterprise.html

Zadeh, L. A. (1965, June). Fuzzy sets. *Information and Control, 8*(3), pp. 338-353.

Zhong, R. (2018, November 8). At China's Internet conference, a darker side of tech emerges. *The New York Times*.Retrieved from https://www.nytimes.com/2018/11/08/technology/china-world-internet-conference.html

Acknowledgments

The debt I contracted in the writing of this book is so large it can't be repaid, barely acknowledged.

The legendary William Zinsser started me on this journey, though that brilliant writer's writer and I met not through writing, but through music—exploring together over more than a decade the American Songbook. Toward the end of his life, when he'd become too blind to read, I'd guide him from the piano to sit, where I'd read aloud to him: sometimes his own writing, sometimes passages from my manuscript. By then he was in his 90s, but with taste so acute that a single sentence, a phrase, a word, could make him smile in delight, or frown with affront (he'd have blue-penciled the adjective legendary, but so he was, and is). My memories of those sessions are exquisite.

The two men this book is dedicated to, my late husband, Joseph F. Traub, and my lifelong friend and sometimes co-author, Edward Feigenbaum, were the fiercest supporters any writer could wish. Each of them read early versions of the manuscript, corrected and praised with generosity, and urged me on when I was faltering, because they believed that I was telling an essential story about the early days of artificial intelligence. Mary Shaw generously volunteered to read the entire manuscript (for she had lived through those times too) and suggested amendments I thank her for. Susan Buckley read important parts, and made judicious editorial suggestions, as did Paul Newell. When Joe and I were lucky enough to spend a sabbatical semester in Cambridge, Massachusetts, Patrick Henry Winston spent time with me explicating his own work on the centrality of story telling to intelligence, and arranged for me to meet other, younger members of the MIT faculty whose work was shaping the next generation of AI applications. He also invited me to sit in on beginning sessions of the MIT-Harvard Brains, Minds, and Machines seminars, weekly seminars that continue lustily, bringing together neuroscientists, computer scientists, engineers, and others with an interest in brains, minds, and machines, who learn from and argue with each other weekly. These early sessions were deeply valuable to me, and his death in July 2019 is a profound loss—professional and personal.

With Deirdre David, a literary scholar and biographer, I had illuminating conversations about the art of biography, the art of memoir, and the glories of the 19th century English novel, her professional specialties and my amateur delights. She was especially helpful in enlarging my knowledge of C. P. Snow and his wife, novelist Pamela Hansford Johnson.

A number of AI researchers also took time to be interviewed, including Oren Etzioni, Edward Feigenbaum, Eric Horvitz, Jaron Lanier, Kathy McKeown, Marvin Minsky, Peter Norvig, Tomaso Poggio, Raj Reddy, Dan Siewiorek, Manuela Veloso, Jeanette Wing, Patrick Henry Winston, and many participants in the 2014 AI Summit in Brooklyn, New York. For the digital humanities, I was happy to be able to talk to David Blei, Edmund Campion, Anthony

Cascardi, Charles Faulhaber, Elizabeth Honig, and Niek Veldhuis. Larry Rasmussen was a splendid tutor in ethics.

A mind must inhabit some kind of body, so it's only right that I thank the people who found me in a grave health emergency, my sister, Sandra McCorduck-Marona, and my brother, John McCorduck, and took the first steps to save my life. My accomplished and humane physicians at the New York Presbyterian/Columbia University Hospital, Jerry Gliklich and Peter Green, made sure I'd live to write many another day. Hila Paldi and Lisa Goldin, accomplished Pilates instructors, kept the pre-emergency and the recovered flesh and bones humming. My longtime housekeeper, Beryl Sibblies, lovingly lifted many responsibilities from me, when I know she would have preferred to retire.

An outstanding team at Carnegie Mellon's ETC Press: Signature, including Keith Webster, Brad King, Drew Davidson and Julia Corrin, encouraged and supported me in every way. Rebecca Huehls did a masterful job of fact-checking and copy-editing under the most trying circumstances. I owe each of them my deep thanks.

Pamela McCorduck
July 2019

About the Author

Pamela McCorduck has published ten books, both fiction and nonfiction. Her *Machines Who Think*(1979, 2004) was the first modern history of artificial intelligence, and her later book, *The Fifth Generation*(1983) co-authored with Edward Feigenbaum, alerted the world to the practical reality of AI. For half a century she was married to the late Joseph F. Traub, the second chair of Computer Science at Carnegie Mellon, who died in 2015. She now lives in the San Francisco Bay Area.

About the ETC Press

The ETC Press was founded in 2005 under the direction of Dr. Drew Davidson, the Director of Carnegie Mellon University's Entertainment Technology Center (ETC), as an open access, digital-first publishing house.

What does all that mean?

The ETC Press publishes three types of work:peer-reviewed work (research-based books, textbooks, academic journals, conference proceedings), general audience work (trade nonfiction, singles, Well Played singles), and research and white papers

The common tie for all of these is a focus on issues related to entertainment technologies as they are applied across a variety of fields.

Our authors come from a range of backgrounds. Some are traditional academics. Some are practitioners. And some work in between. What ties them all together is their ability to write about the impact of emerging technologies and its significance in society.

To distinguish our books, the ETC Press has five imprints:

- **ETC Press:** our traditional academic and peer-reviewed publications;
- **ETC Press: Single:** our short "why it matters" books that are roughly 8,000-25,000 words;
- **ETC Press: Signature:** our special projects, trade books, and other curated works that exemplify the best work being done;
- **ETC Press: Report:** our white papers and reports produced by practitioners or academic researchers working in conjunction with partners; and
- **ETC Press: Student:** our work with undergraduate and graduate students

In keeping with that mission, the ETC Press uses emerging technologies to design all of our books and Lulu, an on-demand publisher, to distribute our e-books and print books through all the major retail chains, such as Amazon, Barnes & Noble, Kobo, and Apple, and we work with The Game Crafter to produce tabletop games.

We don't carry an inventory ourselves. Instead, each print book is created when somebody buys a copy.

Since the ETC Press is an open-access publisher, every book, journal, and proceeding is available as a free download. We're most interested in the sharing and spreading of ideas. We also have an agreement with the Association for Computing Machinery (ACM) to list ETC Press publications in the ACM Digital Library.

Authors retain ownership of their intellectual property. We release all of our books, journals, and proceedings under one of two Creative Commons licenses:

- **Attribution-NoDerivativeWorks-NonCommercial:** This license allows for published works to remain intact, but versions can be created; or
- **Attribution-NonCommercial-ShareAlike:** This license allows for authors to retain editorial control of their creations while also encouraging readers to collaboratively rewrite content.

This is definitely an experiment in the notion of publishing, and we invite people to participate. We are exploring what it means to "publish" across multiple media and multiple versions. We believe this is the future of publication, bridging virtual and physical media with fluid versions of publications as well as enabling the creative blurring of what constitutes reading and writing.