The Coevolution

The Entwined Futures of Humans and Machines

Edward Ashford Lee

The Coevolution

The Entwined Futures of Humans and Machines

Edward Ashford Lee

PREPRINT FOR REVIEW
Version 1.0
To be published by MIT Press, Spring 2020.

14 Coevolution

Chickens and Eggs

Richard Dawkins famously said that a chicken is an egg's way of making another egg. Is a human a computer's way of making another computer? The machines of this book, if we view them as living beings, are creatures defined by software, not DNA, and made of silicon and metal, not organic molecules. Some are simple, with a genetic code of a few thousand bits, and some are extremely complex. Most live short lives, sometimes less than a second, while others live for months or years. Some even have prospects for immortality, prospects better than any organic being. And they are evolving very, very fast. It is not just technology that is changing. We humans are also changing very fast compared to anything found in nature. The way our society works, the way we think, the way we communicate, and, increasingly, even our biology are all in flux.

Are we humans in control of this evolution? Are we truly the masters of the machines? A naive view is what we might call "digital creationism." In this view, we humans use our intelligence to engineer machines in a top-down fashion, like God. A more realistic view is that we are the sources of mutation in a Darwinian coevolution. The mutations we introduce are not entirely random, but in a modern view of evolution, neither are the biological mutations introduced by nature. Much like what happens in nature, most of the technological mutants we bring forth go quickly extinct, while a few grow to occupy a niche, at least for a while, in a continually changing ecosystem.

We have come to depend deeply on technology, to the point that we would not exist without technology. More precisely, we would perhaps exist, but in far fewer numbers and in a form that most of us would find alien and incomprehensible. The "we" of today is a mashup of technology, biology, and culture.

Of the three elements of this mashup, it is technology that is changing fastest. Rapid change can cause problems. Just as disruptions of our gut biome can make us sick, so can technology disruptions. But just as interventions like probiotics can make us healthier, so can technology.

It is natural to fear change. Is artificial intelligence really an existential threat to humanity? Are we destined to be annihilated by a superintelligent new life form on the planet? Are we destined to fuse with technology to become cyborgs with brain implants that define a new form of quasi-human intelligence? Will we lose control of our machines? If technology is coevolving with humans, then we never really had control. The best we can do is prod the process toward a mutually beneficial symbiosis and deal with the unexpected problems as they arise. Even if we are successful, our future symbiotic selves may not resemble humans of today as much as those who fear change might like.

There are risks, and they are not small. Rapid coevolution is inherently unpredictable, and pathologies are likely to emerge. But we should treat these as pathologies, not as a war with invading aliens. The biggest threat to humanity may not be that the machines will make us irrelevant, but rather that the machines will change the very essence of our being, what it means to be human.

Change is scary, but I, for one, am not nostalgic for any human epoch earlier than now. My lifetime has coincided with the most prosperous and relatively peaceful era in all of human history, and no small part of the credit for that goes to technology. We are not living in Eden, of course, but in many dimensions, the human condition has improved. This does not mean that everything will *continue* to improve, but the better we understand the dynamics of our coevolution with technology, the more likely that it will.

Thinkers such as Vinge, Kurzweil, Bostrom, and Tegmark have written about a runaway feedback loop, where the machines design their own successors, breaking free of any symbiosis with humans. I think they may have overestimated what digital computation can do, but even if they haven't, a more likely outcome is much more powerful (and potentially far scarier) human-machine partnerships. Digital computation is the most potent invention humans have ever come up with, and humans have a horrific track record of

using our inventions to perpetrate atrocities on one another. We humans are the scarier part of this partnership.

A Fourth Age

Kevin Laland, the evolutionary biologist who appeared in chapters 3 and 13, in his 2017 book, *Darwin's Unfinished Symphony*, identifies three distinct ages in the evolution of humankind, genetic, genetic-cultural, and cultural evolution. Perhaps we have entered a fourth distinct age, one that we might call the "synthetic age."

Laland's first age, genetic evolution, is shared with all other living residents of our planet. It is dominated by biology and by the happenstances of the environment. This phase was, until recently, thought to be dominated by a form of neo-Darwinian evolution where random mutation provides diversity, and environmental and competitive pressures weed out those less able to survive and procreate. As we will see, the story appears to be more complicated in ways that make the evolution of machines look more like that of early biological life.

A key feature of Laland's genetic evolution age is that the creatures evolving are buffeted by environmental events that are entirely out of their own control. A dramatic example, first suggested in 1980 by the father-and-son team of scientists Luis and Walter Alvarez, is the Cretaceous-Paleogene extinction event, where an asteroid or comet strike is believed to have wiped out the dinosaurs and many other species approximately sixty-six million years ago.

The second age, genetic-cultural coevolution, Laland estimates, began some four million years ago and accelerated quite dramatically over the last forty thousand years or so. In this age, humanoids and then humans began to have a strong enough effect on their own living environment that a feedback pattern emerged, where the environment affected the genes (as before), and the genes affected the environment (new to this age). Laland argues that this feedback was enabled by the emergence of culture, which he defines as "the extensive accumulation of shared, learned knowledge, and iterative improvements in technology over time."[1].

In this second age, the switch from a hunter-gatherer society to an agrarian society enabled population growth and demanded social organization. This accelerated the evolution, according to Laland:

[1] Laland, *Darwin's Unfinished Symphony*, p. 6 (2017).

> Once population size reached a critical threshold, such that small bands of hunter-gatherers were more likely to come into contact with each other and exchange goods and knowledge, then cultural information was less likely to be lost, and knowledge and skills could start to accumulate.[2]

The key feature of Laland's second age is the effect that humans have had on their own environment, becoming "ecosystem engineers." The Dutch evolutionary biologist Menno Schilthuizen, in his book *Darwin Comes to Town*, points out that humans are not nature's first ecosystem engineers. Earlier examples include ants and beavers, who also altered the ecology in ways that then affected their own development. Such feedback loops are fairly common in nature.

Laland describes the third age as follows:

> Now we live in the third age, where cultural evolution dominates. Cultural practices provide humanity with adaptive challenges, but these are then solved through further cultural activity, before biological evolution gets moving. Our culture hasn't stopped biological evolution—that would be impossible—but it has left it trailing in its wake.[3]

Why is cultural evolution so much faster than biological evolution? It must be because humans are able to be more intelligent in bringing about mutations. As Turing himself said in 1950,

> The survival of the fittest is a slow method for measuring advantages. The experimenter, by the exercise of intelligence, should be able to speed it up.[4]

But Turing was not referring to cultural evolution. He was already referring to what I am calling the fourth "synthetic" age, characterized by what is effectively the emergence of a new life form, one based on silicon rather than carbon. This qualifies as a fourth age because, unlike cultural evolution, where the intelligence is applied to evolve itself, in the synthetic age, human intelligence is being applied to evolve symbionts, the machines, and human intelligence, in turn, is evolving as the machines increase their capabilities. The machine symbionts may later become able to harness their own intelligence to evolve

[2] Ibid., p. 150 (2017).
[3] Ibid., p. 234 (2017).
[4] Turing, "Computing Machinery and Intelligence," (1950).

themselves without interaction with humans, as predicted by Bostrom and the others, but this has not really happened yet. If and when it does, we will have entered a fifth age.

All oversimplifications of history that divide it cleanly into distinct phases are flawed, of course. The boundaries between phases are far from clear. But Laland's ages help us distinguish the mechanisms that drive change. The mechanisms driving biological change are clearly not the same as those driving cultural evolution. We could ask whether we should even be using the same word, "evolution," for both. How closely related to Darwin's original idea are these mechanisms? How good is the analogy when we apply the word "evolution" to the development of machines? As it turns out, even in biology, the meaning of the word "evolution" is evolving. But the stalwart constant that sticks with us since Darwin's time is the principle of natural selection, which applies across all these ages.

Evolution Isn't So Simple After All

Darwin's theory of evolution has, like most scientific theories, evolved with time. Darwin developed his theory long before DNA was understood, so the mechanisms of inheritance and mutation were mysterious. Darwin took for granted, for example, that characteristics an organism acquired during its life could be inherited by its offspring, a view known as Lamarckian inheritance, somewhat unfairly named after the French biologist Jean-Baptiste Lamarck.

The writer David Quammen, in his book, *The Tangled Tree: A Radical New History of Life*, engagingly tells the story of how the theory of evolution has itself evolved. He describes Lamarck as "France's great early evolutionist" who became a bit of a laughing stock for his view of inheritance of acquired characteristics. In Quammen's words,

> The most familiar example of such inherited adjustments, which Lamarck himself offered, is the giraffe. The proto-giraffe on the dry plains of Africa stretches to reach high foliage, its neck lengthens (supposedly) from the effort, its front legs lengthen too, and therefore (again supposedly) its offspring are born with longer necks and front legs. Lamarckism, in that cartoonish form, has been easy to despise but harder to kill off entirely.[5]

[5] Quammen, *The Tangled Tree*, (2018).

As it turns out, this idea has been "harder to kill off entirely" because it is partly true, although probably not for the giraffe example. Several mechanisms contribute to inherited characteristics, including the passing from generation to generation of the genome of symbiotic microbes (the hologenome), adaptations of the immune system, and epigenetics, particularly proteins bundled with chromosomes that affect gene expression. All of these reflect acquired characteristics and add information to what is passed from generation to generation. In chapter 8, I pointed out that DNA carries nowhere near enough information to create a human, so other mechanisms must exist.

The hologenome has an obvious analogy with the coevolution of humans and machines. When we put an iPad in the hands of our two-year-old kids, they "inherit" ways of interacting with their environment such as swiping the screen and pinch-to-zoom, that are encoded in the "genome" ("codome"?) of our symbiotic machines. These mechanisms, and many more, have integrated with our brains and shape our thinking much more than we realize. If these are "mutations" in humans, they are not encoded in our genome and they have certainly not come about from the classic neo-Darwinian mechanism of random mutation followed by natural selection. Random mutation certainly plays a role, but it is not even close to the whole story.

Beyond Random Mutation

Some aspects of relatively new evolutionary theories resemble what we see happening with machines more closely than random mutation. One of the radical new discoveries that Quammen documents is called horizontal gene transfer (HGT). In his words:

> The tree of life is more tangled. Genes don't move just vertically. They can also pass laterally across species boundaries, across wider gaps, even between different kingdoms of life, and some have come sideways into our own lineage—the primate lineage—from unsuspected, nonprimate sources. It's the genetic equivalent of a blood transfusion or (different metaphor, preferred by some scientists) an infection that transforms identity. "Infective heredity."[6]

In Darwin's "tree of life" (see figure 14.1), species fork into subspecies through relatively slow accumulation of small random mutations followed by "survival of the fittest," where

[6] Ibid.

"fittest" means most likely to procreate. We have already seen in chapter 9 that this tree is not so simple in that branches can remerge through hybridization. But it turns out to be even more complicated than that.

Horizontal gene transfer is apparently common in bacteria and leads to much faster evolution than random mutations. It is now understood to be the primary mechanism for the spread of antibiotic resistance in bacteria. Many biologists assume that HGT played a major role in the early development of life, but there is also evidence that it played a role later in much more advanced life forms, including humans. According to Quammen, researchers have identified about one percent of the human genome that very likely got itself inserted through HGT mechanisms in the last few million years.

Bacterial Sex

At least three mechanisms for HGT have so far been identified. They are called *transformation*, *conjugation*, and *transduction*. The first to be discovered, transformation, dates back at least to the 1920s, when a British physician, Fred Griffith, noticed that a harmless bacterium could change suddenly into a virulent form that would cause pneumonia, a leading cause of death in those days.

Much later, in the 1940s, the biologist Oswald Avery, working at the Rockefeller Institute in New York, identified DNA as the material from which genes and chromosomes are made, and he found that free-floating DNA from dead bacteria could lead to the kinds of transformations that Griffith had observed. Live bacteria absorb the dead genetic material through their cell membrane and edit it into their own DNA. The transformation mechanism that Avery identified, which was later called "infective heredity," turned out to be fairly common in bacteria. Keep in mind that this was nearly ten years before the landmark publication in 1953 of Watson and Crick's paper describing the double-helix structure of DNA. Avery was repeatedly nominated for the Nobel Prize, which he never received.

In 1946, a twenty-one-year-old researcher, Joshua Lederberg, took a leave of absence from medical school at Columbia to work at Yale University under the guidance of microbiologist Edward Tatum. At Yale, in less than two years, he met and married another student of Tatum's, Esther Miriam Zimmer, identified a second HGT mechanism that he called conjugation, coauthored with Tatum a paper published in *Nature* on conjugation, wrote and filed a PhD thesis, and accepted an assistant professorship in genetics at the University of Wisconsin at Madison. That was a busy two years.

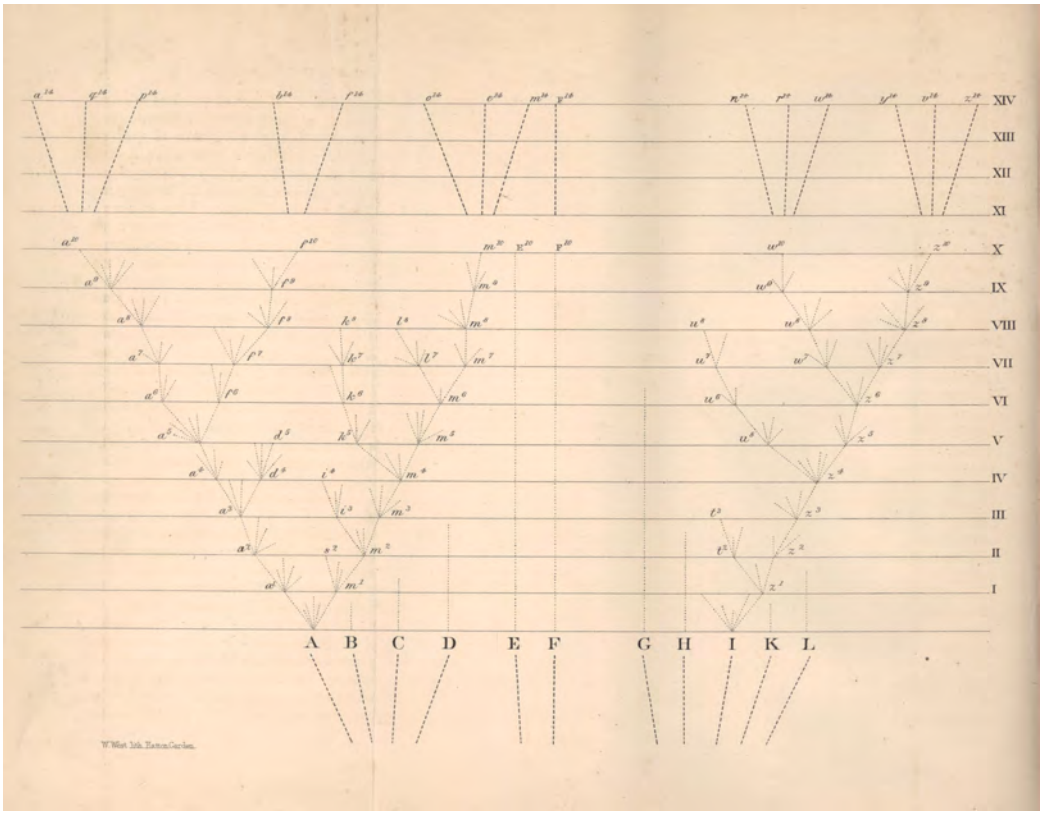

Figure 14.1: The only illustration in Darwin's 1859 *On the Origin of Species* was this depiction of the tree of life. The *A* through *L* at the bottom are hypothetical unnamed species within some hypothetical genus. The lines on the vertical axis labeled I-XIV each represent a thousand generations. The branching shows variation leading to both extinction and new species.

In the mechanism that had been identified by Avery, transformation, a bacterium takes up genetic material left behind by other bacteria that have died. At Yale, Lederberg showed that gene transfer could occur between living bacteria as well. He was not quite twenty-two when his paper with Tatum appeared in *Nature* showing that temporary cell fusion and exchange of genetic material must be occurring. They dubbed the process conjugation and called it a "sexual process."

In 1951, working with his graduate student Norton Zinder in Madison, Lederberg identified a third HGT mechanism that they called transduction, where viruses carry DNA from one strain of bacteria into another. He and his wife, Esther Lederberg, who received her PhD from the University of Wisconsin in 1950, later collaborated to identify a specialized version of transduction that is less random. In 1958, at age thirty-three, Joshua Lederberg shared the Nobel Prize with Edward Tatum and George Beadle for his work on genetics. In that year, he moved to Stanford, where he founded the Department of Genetics.

Lederberg later made significant contributions to computer science. In the 1960s, he played a central role in the development at Stanford of an influential AI program called Dendral, which helped organic chemists to identify unknown organic molecules. This was a GOFAI-style expert system based on encodings of knowledge of chemistry in the form of production rules (see chapter 4).

Horizontal Code Transfer

Horizontal gene transfer upended evolutionary biology. The mechanisms identified by Avery and Lederberg result in faster evolution than the neo-Darwinian mechanism of random mutation followed by natural selection. What biologists mean by "random mutation" here is mutation that is caused by extraneous factors, not by any process that is part of the normal biological processes of the organism. They can occur, for example, due to x-rays or environmental toxins. Although we have seen in chapter 11 that the notions of randomness and causation are far from simple, some sources of mutations, such as x-rays, are clearly extraneous. For some time, many biologists believed that such extraneous mutations were the dominant source of variation in evolution.

The vast majority of such random mutations, however, affect cells that are not germline cells, those involved in procreation, such as eggs and sperm, and hence do not get passed on to offspring. Moreover, the vast majority of random mutations are deleterious and therefore will also not be passed along. If this were the only source of variation, then evolution would likely be much slower than it is.

HGT provides faster random mutations, but it also creates opportunities for less random, more targeted mutations, where strains with a beneficial gene can transfer those genes to entirely different strains of bacteria. This discovery shook the foundations, calling into question the concept of a species and the tree of life, at least with respect to bacteria.

Later discoveries show that HGT occurs throughout nature, not just in bacteria. Human DNA contains significant segments that appear to come from bacteria and even viruses. Carl Zimmer, (2010) reports in *The New York Times* that scientists have found some 100,000 elements in the human DNA that probably come from viruses.

It is tempting to draw an analogy to computer viruses, such as those that are picked up by opening documents sent in a phishing email message. A notable example is the Melissa virus, which infected Microsoft Word. You receive an email from a friend with a message like "Here is that document you asked for; don't show it to anybody else." You open the file, and a macro embedded in the file accesses your contacts in Microsoft Outlook and sends those people email messages similar to the one you just received. Is this analogous to transformation, where Microsoft Word is absorbing "genetic" material (in the form of macros) from its environment and splicing it into its own "genome"? This is not a very good analogy, however. The infected Microsoft Word does not pass along the mutation to any offspring, but rather passes it along to peer individuals. The mechanism is more analogous to the spreading of a disease like the common cold.

A much better analogy can be found in the software development process itself. Software engineering is the discipline of creating new strains of digital machines. The code that is engineered is the "DNA" of the machines. But a software engineer rarely starts from scratch. It is much more common to start with a working program and modify it. A software engineer, therefore, is the source of mutation in a neo-Darwinian "tree of code."

But like the tree of life, the tree of code gets tangled. A software engineer will frequently splice into the code of one program fragments from another. The engineer is acting like the viruses in transduction that carry DNA from one cell to another. We can call this process "horizontal code transfer."

Analogous to transformation, the first HGT mechanism, an engineer will pick up code fragments that are not living, in that they are not part of working programs but are found on the Internet or in libraries of software components. The engineer will splice those fragments into a new program. Beneficial components, ones that have proven useful in many programs, are more likely to get spliced into the "codome" of a new program.

Engineering, like evolution, is about creating artifacts and processes that have never before existed. We engineers tend to think of our role in this process as that of creator, an intelligent designer who, in a top-down fashion, coerces matter and energy to do our bidding. We take great pride in the outcome, our creation, our invention, like our children. But also like our children, we have less control over the outcome than we imagine. Kevin Kelly, whom we met in chapter 2, cites what he calls the "adhocracy" of Wikipedia as evidence that we don't need much top-down design to get fantastic outcomes.[7] We only need a little. We are arguably more like mediators of mutation in an evolutionary process than top-down intelligent designers.

Moreover, our own thinking, during the process of engineering software, evolves along with the machines we build. The software supporting software development shapes the process, splicing memetic material into our cognitive "genome" ("memome"?). The software tools we use to create software change our minds, which in turn changes the software we write.

Top-Down Intelligent Design?

The philosopher Daniel Dennett, in his 2017 book, *From Bacteria to Bach and Back*, makes the case that the human mind, our consciousness, languages, and cultures, are the result of an evolutionary process. He is not talking about the brain and its biological structure and processes, but rather is saying that the mind emerges from more than biology. Dennett is defending and elaborating on the earlier controversial position famously put forth by Richard Dawkins in his 1976 book, *The Selfish Gene*, where Dawkins coined the term "memes" for cultural artifacts and ideas, drawing an analogy between their propagation in human culture and Darwinian evolution. From Dawkins:

> I think that a new kind of replicator has recently emerged on this very planet. It is staring us in the face. It is still in its infancy, still drifting clumsily in its primeval soup, but already it is achieving evolutionary change at a rate which leaves the old gene panting far behind. ... The new soup is the soup of human culture. We need a name for the new replicator, a noun which conveys the idea of a unit of cultural transmission, or a unit of imitation.[8]

That noun is "meme."

[7] Kelly, *The Inevitable*, (2016).
[8] Dawkins, *The Selfish Gene*, (1976).

Dawkins had quite a few detractors who did not like his analogy with biology, but Dennett argues that even some of the most fervent detractors espoused, using other words, essentially the same theory that ideas, culture, and languages propagate via a neo-Darwinian natural selection, where, in Dennett's words, "fitness means procreative prowess." The post-neo-Darwinian mechanism of horizontal gene transfer may be an even better analogy because mutation of ideas is not, mostly, randomly caused by factors entirely extraneous to culture.

Drawing an analogy between biological evolution and evolution of machines is easier even than between biological evolution and evolution of memes because digital machines are more like biological living beings than memes are (see chapter 2). Memes do not have an autonomous existence in the physical world, independent of human brains, but machines do.

Dennett, however, falls short of identifying today's technology as part of his evolving ecosystem. On the contrary, he points to digital technology and software as a canonical example of an opposite kind of design from evolution, what he calls "top-down intelligent design." I will shorten this to "TDID" to avoid repeating the phrase too often. Dennett argues that TDID is less effective than evolution at producing complex behaviors, contrary to a religious position held by some that the complexity of life proves the existence of God.

Dennett points to an elevator controller, observing that every contingency, every reaction, every behavior of the system is imposed on it by the cognitive engineer who designed it. This is partly true, but many aspects of the design have been heavily influenced by prior technology developments,so even a modest elevator controller is the result of an evolutionary process. Moreover as machines go, an elevator controller is a rather simple one. For more complex digital and computational behaviors, like those in Wikipedia, a banking system, or a smartphone, it is hard to identify any cognitive being that performed anything resembling TDID. These systems evolved through the combination of many components, themselves similarly evolved, with engineers introducing both mutations and horizontal code transfer. Together with decades-long iterative design revisions with many failures along the way, we get a Darwinian process of mutation and natural selection.

Dennett argues that, unlike biological beings, the parts of a digital design have no yearnings for resources, nothing driving them forward, no purposes or reasons, and that they are just reactive automata. But this isn't a useful distinction because many alternative designs and mechanisms died along the way, and the ones that survived did so for Darwinian reasons, because they were able to propagate. Barring unsound teleology, propagation is

also the closest that biological evolution gets to having a purpose. The propagation of machines is facilitated by the very concrete benefits they afford to the humans that use them, for example by providing those humans with income and hence with food and the ability to procreate.

Viewing software as "top-down intelligent design" falls victim to the same tendency that Dennett criticizes, the homunculus in the brain, a little man or committee that observes and drives the decision making of the human mind. In contrast, a coevolutionary stance says that software evolves in much the same way that bacteria evolve, through a goal-less coevolution with humans driven by its own Darwinian reward functions, survival and propagation. The tendency to see these designs as TDID is anthropocentric, a tendency that we, as humans, find naturally difficult to avoid. We do not like seeing our mental cognitive processes themselves as cogs in a relentless purposeless evolution. But is this what they are?

Facilitators or Inventors?

Dennett even applies his top-down intelligent design principle to artifacts that are far too complex to have been designed this way:

> To take the obvious recent example of such a phenomenon, the Internet is a very complex and costly artifact, intelligently designed and built for a most practical or vital purpose: today's Internet is the direct descendant of the Arpanet, funded by ARPA (now DARPA, the Defense Advanced Research Projects Agency), created by the Pentagon in 1958 in response to the Russians beating the United States into space with its Sputnik satellite, and its purpose was to facilitate the R&D of military technology.

This is an oversimplification of the Internet. ARPA funded the development of a few of the protocols that underlie the Internet, but even these protocols emerged from many failed experiments at methods for getting computers to interact with one another (see chapter 6 of Lee, 2017). Moreover, ARPA and DARPA had little to do with most of what we recognize as the Internet today, including web pages, search engines, YouTube, and so on. As I pointed out in the previous chapter, Tim Berners-Lee, creator of the web, laments what it has become. Much of the Internet evolved from the highly competitive entrepreneurial dog-eat-dog ecosystem of Silicon Valley and the collaborative minds of thousands of contributors to the standards that make it as robust as it is today.

The computer scientist and entrepreneur Danny Hillis, referring to the Internet, writes

> Although we created it, we did not exactly design it. It evolved. Our relationship to it is similar to our relationship to our biological ecosystem. We are codependent and not entirely in control.[9]

Further digging himself in, Dennett demurs,

> All of this computer R&D has been top-down intelligent design, of course, with extensive analysis of the problem spaces, the acoustics, optics, and other relevant aspects of the physics involved, and guided by explicit applications of cost-benefit analysis, but it still has uncovered many of the same paths to good design blindly located by bottom-up Darwinian design over longer periods of time.[10]

Dennett does not see that "computer R&D" is actually more like Dawkin's memes than like TDID. Humans are more facilitators than inventors, and as Dennett notes about culture,

> Some of the marvels of culture can be attributed to the genius of their inventors, but much less than is commonly imagined ...

The same is true of technology.

Evolutionary Multipliers

Although Dennett overstates the *amount* of top-down intelligent design in technology, there can be no doubt that human cognitive decision making strongly influences its evolution. At the hand of a human with a keyboard, software emerges that defines how a new machine strain reacts to stimulus around it, and if those reactions are not beneficial to humans, the strain very likely dies out. But this design is constructed in a context that has evolved. It uses a human-designed programming language that has survived a Darwinian evolution and encodes a way of thinking. It puts together pieces of software created

[9] Hillis, "Introduction: The Dawn of Entanglement," (2011).
[10] Dennett, *From Bacteria to Bach and Back*, (2017).

and modified over years by others and codified in libraries of software components. The human is partly doing design and partly doing random mutation, horizontal code transfer, and simple husbandry, "facilitating sex between software beings by recombining and mutating programs into new ones."[11] So it seems that what we have is a *facilitated* evolution, facilitated by elements of top-down intelligent design and conscious deliberate husbandry. There are many examples of facilitated evolution in nature, including, for example, the Cambrian explosion (see chapter 2); human husbandry of farm animals, domestic pets, and crops;[12] evolution of animals and plants to adapt to human urbanization;[13] and the development of antibiotic-resistant bacteria via horizontal gene transfer.

As with any evolutionary process, competition for resources plays a role in the evolution of machines, and death and extinction are natural parts of the process. The success of Silicon Valley depends on failure of startup companies as much as it depends on their success. Software competes for a limited resource, the attention and nurturing of humans that is required for the software to survive and propagate. Consider the browser wars of the 1990s, where many attempts at programs for viewing content on the Internet succumbed to competition in acts of deliberate and systematic killing. Having been caught by surprise by the emergence of the web, starting around 1995, Microsoft built Internet Explorer into all Windows systems, free of charge, in a deliberate attempt to kill off the competing browsers. Today, few browser species survive.

Wikipedia and Google are spectacular multipliers of our human cognitive abilities, but they are not themselves top-down intelligent designs. Although their evolution has most certainly been facilitated by various small acts of TDID, they far exceed as affordances anything that any human could have possibly designed. They have coevolved with their human symbionts.

Dennett observes that collaborating humans vastly surpass the capabilities of any individual human. Humans collaborating *with technology* further multiplies this effect. Technology itself now occupies a niche in our (cultural) evolutionary ecosystem. It is still relatively primitive compared to humans, much like our gut bacteria, which facilitate digestion. Technology facilitates thinking.

[11] Lee, *Plato and the Nerd*, chapter 9 (2017).

[12] For a beautifully illustrated book documenting the effect of humans on the transformation of animals, see Grouw, (2008).

[13] For examples of rapid Darwinian evolution of animals and plants in urban landscapes, see Schilthuizen, (2018).

Parasitic or Symbiotic?

Dennett takes on AI and most particularly deep learning systems, calling them "parasitic." He focuses on their mechanics, noting that while they can classify images, for example, those images have no meaning to them. They "parasitically" derive any meaning from humans. In his words,

> Deep learning (so far) discriminates but doesn't notice. That is, the flood of data that a system takes in does not have relevance for the system except as more "food" to "digest."[14]

This limitation evaporates when these systems are viewed as symbiotic rather than parasitic. In Dennett's own words, "deep learning machines are dependent on human understanding."

Dennett notices a similar partnership between memes and the neurons in the brain:

> There is not just coevolution between memes and genes; there is codependence between our minds' top-down reasoning abilities and the bottom-up uncomprehending talents of our animal brains.[15]

For the neurons in our brain, the flood of data they experience also has no "relevance for the system except as more 'food' to 'digest."' An AI that requires a human to give semantics to its outputs[16] is performing a function much like the neurons in our brain, which also, by themselves, individually have nothing like comprehension. It is an IA, intelligence augmentation, not an AI.

Dumbing Down

Today, there is a lot of hand wringing and angst about AI. Dennett raises one common question:

> How concerned should we be that we are dumbing ourselves down by our growing reliance on intelligent machines?[17]

[14] Dennett, *From Bacteria to Bach and Back*, (2017).
[15] Ibid.
[16] Lee, *Plato and the Nerd*, chapter 9 (2017).
[17] Dennett, *From Bacteria to Bach and Back*, (2017).

Are we dumbing ourselves down? It doesn't look that way to me. This does not mean we are out of danger. Far from it. Again, from Dennett:

> The real danger, I think, is not that machines more intelligent than we are will usurp our role as captains of our destinies, but that we will overestimate the comprehension of our latest thinking tools, prematurely ceding authority to them far beyond their competence.

I believe there are far bigger dangers than this one. First, IA in the hands of nefarious humans and governments is a scary prospect indeed. Second is that the machines will change our thinking, as they already have, through the creation of filter bubbles and echo chambers.

Evolutionary pressures may tend to accentuate the fragmentation of information through a phenomenon that evolutionary biologists call the Baldwin effect, named after the American philosopher James Mark Baldwin (1861–1934). Under this effect, an organism's ability to learn new behaviors during its lifetime affects its reproductive success and will therefore have an effect on the genetic makeup of its species through natural selection. Today, a search engine that acquires enough "knowledge" of me to tune its results to what I want to hear is more likely to survive and propagate in the ecosystem of search engines, which compete for advertising dollars. As the search engine learns, its reproductive prowess improves, thereby reinforcing the development of machines that fragment human thinking. As it learns, it creates for me an ever smaller echo chamber, feeding me only the information I want to see. And its progeny will even more effectively isolate our progeny from each other.

A third danger bigger than the one Dennett cites is that the machines will shed their dependence on humans and that we will lose control. This is the danger that Bostrom, Tegmark, and others focus on. It is true that there have been moderately successful experiments where programs learn to write programs, and it seems inevitable that the machines will continue to get better at designing themselves. This fear is real, but it may be that we never really were in control, so losing control is not the essential issue. If the machines are evolving in a Darwinian way, then the best we can do is nudge the process. We cannot really control it, but through policy and regulation, we may be able to slow or even prevent undesirable outcomes.

Dennett's final words are optimistic:

> If our future follows the trajectory of our past—something that is partly in our control—our artificial intelligences will continue to be dependent on us even as we become more warily dependent on them.

I share this optimism, but also recognize that rapid coevolution, which is most certainly happening, is extremely dangerous to individuals. Rapid evolution necessarily involves a great deal of death. Both technologies and memes will fall by the wayside as the symbiogenesis evolves. *Co*evolution means that both parties, the humans and the technologies, will change. Even if this remains symbiotic, the results can be dramatic. The resulting humans may be very different from the humans of today.

Endosymbiosis

Analogies can be useful intuition pumps, to use the words of Dennett, but they are risky. I am drawing an analogy between evolution of digital technology and both biological evolution and Dawkins's memetic evolution. Just as with memes, for digital technology, mutation and natural selection both occur, where humans provide the mechanisms for both. It is not just an analogy. The parallel with life *is* an analogy, but that parallel is not so important. What is important is that we understand the mechanisms of change, and not oversimplify by vilifying individual technologists each time we discover a pathology in the evolving ecosystem. If these mechanisms of change truly were top-down intelligent design, then vilifying the engineers may be justified. But the mechanisms are more complex. We are all complicit, for example, in the shape of the ecosystem that determines whether a technology strain either succeeds and propagates or fails and goes extinct. The engineers act like the viruses in horizontal gene transfer, transporting "genetic" material from one technological strain to another. But the rest of society overuses antibiotics, thereby creating an ecosystem that naturally leads to antibiotic-resistant bacteria.

Although we *are* facing the possibility of the machines affecting our genes, today, the human side of the coevolution is still mostly memetic rather than biological. Our ability to mutate technology, to engineer new strains, evolves like Dawkins's memes, considerably pushed along by the technology itself, which provides the software and hardware that we routinely use to engineer new software and hardware. A strong feedback loop forms, where technology causes memetic mutation, and the memes cause technology mutation.

It turns out, however, that there is an even stronger and scarier analogy with biology. The mutual dependence we have with technology is a symbiosis, and symbiosis can lead to

an even stronger source of mutation than horizontal gene transfer. Biologists call this source of mutation *symbiogenesis*. It is where an entirely new and more complex life form emerges from a fusing of the partners in a symbiosis. Symbiogenesis is also called *endosymbiotic theory*, and an endosymbiosis is a symbiosis where no partner can live without the other—like its weaker cousin, an obligate symbiosis—but the partners have fused to become one, where one lives within the tissues of the other.

The biologists David Smith and Angela Douglas give cows as an example of an endosymbiosis. Cows, they say, are "forty-gallon fermentation tanks on four legs."[18] Lynn Margulis, who deserves much of the credit for our current understanding of symbiogenesis, describes cows this way:

> Cows ingest grass, but they never digest it because they are incapable of cellulose breakdown. Digestion in cows is by microbial symbionts in the rumen. The rumen is a special stomach, really an overgrown esophagus, that has changed over evolutionary time. Cows that lack rumens don't exist; cows (and bulls) deprived of their microbial symbionts are dead.[19]

A cow is not a creature that contains microbial symbionts. Rather, the symbionts are no less part of the cow than the rumen itself. Without the symbionts, there is no cow.

Human dependency on technology has not quite reached this stage, in the sense that humans would continue to exist without technology, albeit in far fewer numbers. But the strength of the codependence keeps increasing, and it is not farfetched that we will reach a point where what we mean by "a human" includes the technologies without which that human cannot live.

Evolutionary Discontinuity

The relationship between a cow and its gut microbes is asymmetric. The microbes are physically much smaller and biologically simpler than the cow. The relationship between humans and technology today is also asymmetric. Digital artifacts are far simpler than our brains, and we at least have the illusion of being in control, using technology as a tool. This asymmetry will likely decrease over time as technology gets more sophisticated,

[18] Smith and Douglas, *The Biology of Symbiosis*, (1987).
[19] Margulis and Sagan, *Acquiring Genomes*, pp. 14-15 (2002).

and the resulting symbiosis could become first an obligate symbiosis and eventually an endosymbiosis.

In biology, there are less asymmetric endosymbioses than that of a cow. The human cells in our bodies, as well as those in all plants and animals, very likely emerged as an endosymbiosis of simpler creatures. These cells are quite different from those of bacteria, which lack mitochondria, chloroplasts, and a nucleus. Those organelles have their own enclosing membranes, and most biologists today believe that they evolved from independent creatures that fused to form today's cells. Biologists call cells with such organelles *eukaryotes* and distinguish them from *prokaryotes*, which, like bacterial cells, have no such internal structure. Eukaryotes evolved from a symbiosis between prokaryotes. The importance of this step cannot be overstated:

> The largest evolutionary discontinuity on this planet is not between animals and plants; it is between prokaryotes (bacteria without membrane-bounded nuclei) and eukaryotes (all the others made of cells with membrane-bounded nuclei). The detailed story of this huge discontinuity is connected to the origins of species.[20]

The evolutionary biologist Ernst Mayr called the emergence of eukaryotes "perhaps the most important and dramatic event in the history of life."[21] The merging of humans with technology, if it happens, will be equally momentous.

Neo-Darwinian evolution of humans may have slowed because we produce fewer offspring than we used to, so there are fewer mutations per parent, and those offspring are more likely to survive and reproduce. Better health care, clean water, and safe food attenuate the effect of natural selection. Put differently, the memetic evolution that keeps us from drinking the water pooled in the gutter affects the gene pool, illustrating the Baldwin effect.

Further evolution of the human genome may, in the future, occur more through genetic engineering than through random mutation or horizontal gene transfer. George Dyson speculates:

> Are we using digital computers to sequence, store, and better replicate our own genetic code, thereby optimizing human beings, or are digital computers

[20] Ibid., p. 141 (2002).
[21] Mayr, *What Evolution Is*, p. 48 (2001).

> optimizing our genetic code—and our way of thinking—so that we can better assist in replicating them?[22]

But even without genetic engineering, humanity may change through a symbiogenesis with technology. Our pacemakers and insulin pumps on the biological side, and our banking, transportation, and communication systems on the cultural side, may be the precursors of symbionts without which some future form of humans will become less able to procreate. It is not hard to imagine, for example, a world in which sex never leads to pregnancy and humans lose the ability to become pregnant that way.

Endosymbiotic theory is relatively young. Lynn Margulis was twenty-nine years old when in 1967 she published "On the Origin of Mitosing Cells" under the name Lynn Sagan (she had married and then divorced the famous science popularizer Carl Sagan). Her title is a clear bow to Darwin's 1859 *On the Origin of Species*. In this paper, she resurrected what many biologists considered to be a wacky idea first advanced by the Russian botanist Konstantin Mereschkowski, who, in the early 1900s, suggested that eukaryotic cells evolved from a symbiosis between distinct prokaryotic cells. It was Margulis who put the theory on a sound biochemical footing. In a highly influential book written later with her son, Dorion Sagan, she writes,

> We believe random mutation is wildly overemphasized as a source of hereditary variation. ... Rather the important transmitted variation that leads to evolutionary novelty comes from the acquisition of genomes. Entire sets of genes, indeed whole organisms each with its own genome, are acquired and incorporated by others. The most common route of genome acquisition, furthermore, is by the process known as symbiogenesis.[23]

The genome of a mitochondria within a human cell is distinctly different from that in the nucleus of the cell. Both sets of genes are inherited, although the mitochondrial genes only from the mother. Mereschkowski and Margulis's hypothesis is that, far in the past, one cell ingested another, and instead of digesting it, hijacked its functions to make it part of a new type of cell.

Some human lives already depend on technologies incorporated into our bodies, for example pacemakers. But a pacemaker is not inherited by offspring. If we reach the point where human newborns are routinely augmented with technological prostheses, or

[22] Dyson, *Turing's Cathedral*, p. 311 (2012).
[23] Margulis and Sagan, *Acquiring Genomes*, p. 11-12 (2002).

where procreation is always mediated by machines, we will have entered a new era for biological life. More dramatically, is it possible that we humans will become the mitochondria of the technium, organelles that perform a vital function in the larger being but that cannot live on their own? Today, this is the stuff of science fiction.

But other scary scenarios loom closer. An endosymbiosis forms with the fusing of two life forms into one. Technology today cannot live without us humans, so although our dependence on it is not absolute, its dependence on us is. Will we become like the gut bacteria of technium, able to live outside the host, but only at the cost of very poor health? Gut bacteria do not fare well on their own. Or worse, will we become the parasites or pathologies of the technium, doomed to be subjugated or even annihilated? We are already seeing "machine medicine" and "machine immune systems" improving, where software self-repairs and AIs expunge malware. What if we humans become tantamount to malware?

Will We Be Eclipsed?

Our dependence on technology has been steadily growing, a trend that seems likely to continue, but technology would go extinct overnight without the help of humans. Is it likely to shed that dependence on us? For this to happen, the machines will need to operate, procreate, and evolve without the help of humans.

In 2018, a team of researchers at the University of Toulouse and the University of York created a program that could write programs to play old Atari video games credibly.[24] Their program generated random mutations and then simulated natural selection. Their technique was itself evolved (via horizontal code transfer) from earlier work that evolved programs to develop certain image processing functions.[25] In principle, these projects and many other fledgling efforts on automatic coding show that if the machines somehow figure out how to keep themselves running without the help of humans, they could evolve their software without the help of humans. Moreover, their evolution would be using a method, natural selection, that is known to be effective at producing very sophisticated beings.

The Atari game-playing programs that emerge from the Toulouse-York evolutionary process, however, are far less effective than programs based on deep learning. The Toulouse-York team admits this, saying that the main advantage of their technique is that

[24] Wilson et al., "Evolving Simple Programs for Playing Atari Games," (2018).
[25] Miller and Thomson, "Cartesian Genetic Programming," (2000).

the resulting programs are more explainable (see chapter 6). The game-playing strategies can be read (by humans) from the evolved programs. Such an advantage, however, is irrelevant if there are no humans demanding explanations.

Evolution is a form of learning. To the extent that there is a distinction between evolution and learning, evolution governs what emerges at birth and learning governs what emerges during life. In biological systems, both forms of acquired capability are passed on to offspring, the first primarily through genetics, and the second primarily through memetics.

In both cases, information that passes from one generation to the next over a noisy channel, according to the Shannon channel capacity theorem (see chapter 8), carries only a finite number of bits. In *machine* learning, versus human learning, a finite number of bits is all there is, at least today, and hence capabilities that a technology acquires during its "life" can be passed on *perfectly* to its offspring. Lamarckian inheritance is a reality for digital technology. For biological creatures, the story is less clear, however, because some information is carried from generation to generation by the "thing in itself," the continuous biological process that is some four billion years old. This information is not limited to a finite number of bits.

Moreover, by the Baldwin effect, the introduction of machine learning into a wider variety of technological artifacts will enhance their procreative prowess. Their ability to learn during life will make them more adaptable to changing environmental conditions, which makes them more likely to survive and propagate. For example, if humans were to decide someday to kill off some strains of technology, only those that can adapt to this hostile environment will survive and propagate. We are already seeing human-created regulations and laws prohibiting certain kinds of technologies, and we are seeing adaptation in technology strains to survive these laws. Some technologies also succumb to pathologies, becoming extinct because their weak security makes them too vulnerable to viruses and worms. The inability to adapt can doom a species. For example, in December 2018, Google announced that they would kill Google+, citing new vulnerabilities to malware that were not worth the cost to fix. Google+ was, apparently, insufficiently adaptive.

Immortality

Today, the state of an executing computer program can be copied, stored, and restored perfectly with astonishingly high confidence. This is possible because the essential properties of the program are digital. Inessential properties, such as the temperature of the

chips running the program, cannot be perfectly copied, but those properties do not define the being. Digital traits can also be perfectly passed on to offspring.

However, many digital technologies are not *completely* digital. Robots, for example, are not robots unless they have a physical presence able to interact with the physical world. Self-driving cars are not self-driving cars unless they have wheels and can move through physical space. The robotics researchers Paul Fitzpatrick, Giorgio Metta, and Lorenzo Natale, in a paper entitled, "Towards long-lived robot genes," lament,

> Robot projects are often evolutionary dead ends, with the software and hardware they produce disappearing without a trace afterwards.[26]

As machines become more embodied (see chapter 7), their inheritance mechanisms will inevitably become less perfect.

In chapter 8, I pointed out that any being that is completely defined by a digital code can, in principle, become immortal. Nature, however, has given us no immortal beings. In fact, evolution does not even *favor* longevity, much less immortality! Peter Goddfrey-Smith, who appeared in chapter 2 with his study of octopuses, has pointed out that every evolutionary advantage comes with a cost, and evolution favors advantages that help early in life, before and during procreation, at the expense of costs incurred later in life, after procreation. This explains why evolution has not and probably never will deliver immortality. As we age, we pay for the strong body we once had. Embodied machines will likely similarly never develop immortality. They too will age and die.

Intellectual Sidelining

Even if embodied robots fail to eclipse humans, to the extent that intelligence can be accomplished in a purely digital way, humans still may be intellectually sidelined. In a 2016 TED talk, Sam Harris, whom we met in chapter 10, made the case that intelligence is information processing, and that the information-processing abilities of our machines will continue to improve. He concludes that it is only a matter of time before they eclipse us. Harris is not alone in drawing such a conclusion. Nick Bostrom, Max Tegmark, and Kevin Kelly have all written similar predictions.

Despite my argument in chapter 1 that intelligence does not lie on a linear scale, the prediction is hard to refute. It is certainly possible for the machines to continue to improve

[26] Fitzpatrick, Metta, and Natale, "Towards Long-lived Robot Genes," (2008).

in *all* relevant dimensions of intelligence, in which case they could certainly sideline us. However, these writers do not make any distinction between digital information and nondigital information. If the latter is essential, then we have not yet invented the technology that will eclipse us.

My argument in chapter 8, that cognition (probably) is not digital and algorithmic, can, perhaps, just slow down our progress toward doom. As we learn more about the neuroscience of intelligence, it will become easier to make machines that *do* include the right sorts of processes to match and exceed any cognitive function in humans. Our brains are irrefutable proof that it is possible to make intelligent machines, since nature has done so. Is it farfetched to assume that only biochemical machines driven by human DNA are capable of such intelligence? Could the concept of embodied cognition save us? Digital machines will never have human bodies.

In chapters 11 and 12, we saw that interaction is more powerful than computation. In chapter 7, we saw that interaction with the physical world is central to cognition. Although, today, machines are far less embodied than humans, they are interacting with the physical world more every day. Kai-Fu Lee, whom we met in chapter 13, points out that China's Internet and AI infrastructure already penetrates deeply into the physical world. Such eyes and ears are the first step toward an embodied cognition.

The second step is for the machines to manipulate the physical world. Just as Facebook's machines can experiment with user interface designs (see chapter 11), learning from the reactions of the users, TenCent's machines can experiment with physical actions. How does placement of bicycle stands affect mobility in a city? How does pricing of services affect where people go? How can users be incentivized to leave scooters where they are most likely to be picked up and used again? As these computer systems close the feedback loop, affecting the physical world and measuring its reaction, will this reafference (see chapter 5) inevitably result in self-awareness and human-like intelligence? My guess, and it is just a guess, is that the intelligence that will emerge will not resemble human intelligence much at all. But this is far from reassuring.

There is another weakness in the argument that humans will be sidelined, although this weakness is also far from reassuring. The doomsday scenarios compare humans of *today* to machines of *tomorrow*. But humans will change too, and indeed we already are changing. Our cognitive and physical beings are already intertwined with the machines, and this co-dependence and integration is only going to accelerate. This does not necessarily result in a less scary picture, however.

Soulless Machines?

Despite the emergence of AI-generated art (see chapter 10), perhaps we can derive solace from a soulful sensation that only humans can possibly create and appreciate poetry, music, and dance. Douglas Hofstadter expresses this sensation this way:

> Many educated people believe that although a machine may now or someday be able to do a creditable job of acting like a person, any machine's performance will always remain lackluster and dull, and that after a while this dullness will always show through. You will simply have no doubt that the machine is unoriginal, that its ideas and thoughts are all being drawn from some storehouse of formulas and clichés, that ultimately there is nothing alive and dynamic—no *élan vital*—behind its façade.[27]

The utterances of Tay on Twitter, however, were anything but dull (see chapter 10).

Reducing art to neuroscience, Steven Pinker says,

> The real medium of artists, whatever their genre, is human mental representations. Oil paint, moving limbs, and printed words cannot penetrate the brain directly. They trigger a cascade of neural events that begin with the sense organs and culminate in thoughts, emotions, and memories.[28]

If the essence of art is "human mental representations," then by definition, machines cannot participate. They are not human. The purpose of art becomes the conveyance of these mental representations from one human to another.

We have a word for words that are especially economical and effective at conveying mental representations. We call these words "poetry." But even a poem is imperfect. The thoughts it triggers in your mind will not match those in the mind of the poet no matter how poetic the words are. Often, the power of poetry lies in its ambiguity and its ability to adapt to the individual, to trigger powerful and personal emotional thoughts in a human whose cognitive world is very different from that of the poet.

We have already seen that technology gives an artist a richer palette and more versatile media. It has never been the case that the art is created by a paintbrush, but a good paintbrush can make a big difference. And the most effective paintbrush is the one designed

[27] Hofstadter, "Can Inspiration Be Mechanized?," (1982).
[28] Pinker, *The Blank Slate*, p. 417 (2002).

to work well with the human hand and human eye. As machines get ever more deeply synergistically intertwined with our human world, they will inevitably provide us with more media for creativity. Their role in the human soul, therefore, is not to replace it with dry objectivity. Instead, they have real potential to enrich our artistic lives by providing us with entirely new kinds of paintbrushes.

Recall from chapter 10 that Plato, in *Timaeus*, asserts that if we understand the mechanisms that cause a human action, that action becomes soulless, one for which we cannot hold the human accountable. When we assume that the actions of a machine will be soulless, it is perhaps because we assume that the mechanisms behind those actions are explainable. But as we saw in chapter 6, modern AI programs yield behaviors we cannot explain. This may be the reason that artists have taken note and are starting to use AI as an art medium.

But we can go even further. Today's software is digital and algorithmic. The physical world, on the other hand, is (probably) neither digital nor algorithmic, and it can exhibit both nondeterminism and chaos, both of which make behaviors fundamentally unpredictable. Future machines may harness both of these to produce genuine delight in their human symbionts.

Ethical Technology

Digital technology today is a tsunami swamping human culture. It is changing our political systems, economies, and social relationships. It is redefining our intellectual lives, changing how we pursue science, anthropology, art, and literature. It creates fabulous wealth and opportunity while devastating entire careers. It informs and misleads, unites and divides, and empowers and paralyzes. It unleashes free speech and enables ubiquitous surveillance. And that is just today. What about tomorrow?

There are enormous opportunities and risks. How can we mitigate the risks and maximize the opportunities? Many educators believe that the answer is to teach ethics in engineering and computer science schools. If this is indeed a solution, then a corollary is that bad outcomes are the result of unethical actions by one or more individuals. But given the complexity of socio-technical interactions and the coevolution thesis of this book, this corollary is probably invalid. It is analogous to the assumption that if each neuron in the brain is operating normally, then mental illness cannot emerge. Under that assumption, mental illness could be treated by identifying the rogue neuron or neurons and killing them. I don't think any credible psychiatrist or neuroscientist is pursuing such a route.

While it is certainly important that engineers behave ethically, teaching ethics is not a panacea. Even if we could get every technology developer to behave ethically, an unrealistic goal, pathologies will still emerge. Many of the detrimental effects that we have seen are unintended and unanticipated consequences of well-meaning actions. If we overemphasize ethics, we could end up just vilifying scapegoats without really improving anything.

Today, the only effective principle guiding technology development seems to be the pursuit of profit. This is a strong motivator, and it stimulates creativity, but it is a blunt instrument, and history has shown that it must be regulated. To do better, we have to first understand the complex dynamics of an evolving socio-technical culture.

Some people use the term "digital humanism" for a human-centric study of technology. It is imperative for intellectuals of all disciplines to step up and take seriously this intellectual challenge. Our limited efforts to rein in the detrimental effects of technology have been, so far, mostly ineffective, underscoring our weak understanding of the problem. The privacy laws in the United States and Europe, for example, are not accomplishing their objectives. And it is not even clear that the privacy goals can be met even if all human participants behave ethically, an unrealistic expectation.

What Should We Teach the Young?

Humans have a handicap compared to machines. Because of the digital nature of their knowledge, everything that a digital machine learns can be copied nearly instantaneously to another similar machine. Humans, on the other hand, start from scratch and have to go through a decades-long painful and imperfect process of knowledge transfer that we call "education." This handicap, however, is also an opportunity. If we start early, focusing young minds on the hard questions of digital humanism, perhaps we have a chance. After all, it is the next generation that will both drive innovation and bear the brunt of the mistakes.

Traditionally, a well-educated person is one with knowledge of language, history, and science, and the skills to manipulate formal systems like mathematics and computer programs. Today, it seems that wisdom takes a back seat to skills and knowledge of facts. This has proved valuable, making our young more employable. Skills and facts, however, are increasingly becoming better handled by machines, so perhaps these are not the best choices for what to emphasize in the future.

It is clear that our young should study technology, but, I believe, not primarily to enhance their job prospects, but rather to enhance their understanding of the society they are growing up in. It is a nice side benefit that, in the short term, it *will* enhance their job prospects, but the durable value comes from developing a deeper understanding of the tectonic forces that will make all those skills obsolete and their knowledge superfluous.

Instead of just teaching kids how to write programs in Python, for example, we should also introduce them to Guido van Rossum at CWI in The Netherlands, the creator of the original Python, and to the open-source community that has grown up around Python. They should develop an understanding of the sociology of Python and open-source software.

We should introduce our young to ideas around privacy, using their own tools—Snapchat, WeChat, Instagram, and Facebook—as illustrations. Privacy is a fascinating philosophical conundrum and a relatively recent concept. Studying the technology around privacy can lend insights into what it really means for humans. We should help them understand the dynamics of viral spread of ideas. This is a very different teaching agenda than teaching them how to write programs that sort numbers, the focus of most introductions to computing today.

Sadly, most educators do not do the sort of teaching I have in mind. For most of my career as a professor, neither did I. It would not have occurred to me that Python has a history, that it emerged from the mind of a single creative individual and then evolved into an entire ecosystem of technological species, most of which will go extinct. To me, Python was a Platonic fact about the world and may as well have always existed. To me, all programming concepts had this character.

I have a very different view today, but I spent many years spreading a profound misunderstanding of technology, one that locks our young into a hopeless acceptance of technological "facts" about the world. We think we are empowering them by giving them the skills to get jobs, but we are actually shutting them in to today's facts and making them vulnerable to a changing world. It is perhaps ironic that the technologists of tomorrow need to be our strongest humanists.

Public Policy

Humans need heroes. We like to single out brilliant individuals and give them Nobel Prizes and credit them as inventors and entrepreneurs. Every time we do this, we ignore thousands of other individuals, each of whom was essential to the outcome. On the flip

side of the coin, when technology leads to bad outcomes, we like to single out individual villains. Hence, we drag Silicon Valley executives in front of Congress and threaten to break up their companies. Assigning blame to greedy capitalists may make us feel good, but it has little effect on future technology outcomes. Attacking capitalism itself will affect future societal outcomes, even if not technology outcomes, but most of us probably will not like those outcomes. The twentieth century tried that experiment. So what should we do to prevent bad technology outcomes?

Under digital creationism, the purpose of regulation is to constrain the individuals who develop technology. Under coevolution, the purpose of regulation is to nudge the process of technology development. Under digital creationism, bad outcomes are the result of unethical actions by individuals, for example by blindly following the profit motive with no concern for societal effects. Under coevolution, bad outcomes are the result of procreative prowess. Technologies that succeed are those that more effectively propagate. The individuals we credit with creating those technologies certainly play a role, but so do the users of the technologies. Should we establish policies to constrain those users?

Consider privacy laws. I believe these have been ineffective because they are based on digital creationism as a principle. These laws erroneously assume that changing the behavior of corporations will be sufficient to achieve privacy goals. A coevolutionary perspective understands that users of technology will choose to give up privacy even if they are explicitly told that their information will be abused. We are repeatedly told exactly that in the fine print of all those privacy policies we don?t read.

I don't have a concrete proposal that will effectively improve personal privacy. I am not even sure what it means to improve personal privacy. I value freedom, so individuals should be free to give up their own personal privacy. Most of the people that I talk to tell me that they have nothing to hide, so they don't mind giving up their privacy. But what if the collective actions of many such individuals leads to an Orwellian state, as it has in China?

I believe that, as a society, we can do better. I'm not sure how to prevent an Orwellian state (or perhaps, worse, a corporate Big Brother). But I am sure that we will not do better until we abandon digital creationism as a principle.