4

THE GLOBAL CHOICE

What happens fast is illusion, what happens slowly is reality. The job of the long view is to penetrate illusion.

• STEWART BRAND

In the early 1960s, during the craze for war games that shaped so much of Cold War military strategy, the Naval War College acquired a $10 million computer. Its purpose was not to calculate torpedo trajectories or to help plan shipbuilding budgets. It was, instead, a game machine known as the Naval Electronic War Simulator. By managing the simulations of war games, the computer could amplify the decision-making powers of the military commanders, since a computer could presumably model a much more complex set of relationships than a bunch of humans rolling dice and moving tokens around a game board. It is unclear whether the Naval Electronic War Simulator actually improved US military decision-making in the years that followed. Certainly, the ultimate path of the Vietnam War suggests that its intelligence amplification was limited at best.

The idea of a computer smart enough to assist with complex decisions may have been premature in the 1960s, but today it no longer seems like science fiction. Ensemble forecasts from meteorological supercomputers help us decide whether to evacuate a coastal area threatened by a hurricane. Cities use urban simulators to evaluate the traffic or economic impact of building new bridges, subways, or highways. The decisions that confounded some of the finest minds of the nineteenth century—the urban planners filling in Collect Pond, Darwin and his water cure—are increasingly being guided by algorithms and virtual worlds.

Supercomputers have started taking on the role that in ancient times belonged to the oracles: they allow us to peer into the future. As that foresight grows more powerful, we rely on these machines more and more to assist us in our hard choices, and perhaps even to make them for us. It's easy enough to imagine computer simulations and forecasts helping to decide the future of Collect Pond: projecting population growth in downtown Manhattan, the ecosystem impact of destroying a freshwater resource, and the economic fortunes of the tanneries polluting that water.

Almost a hundred years ago, when Lewis Fry Richardson alluded in his "Weather Prediction by Numerical Process" essay to the "dream" of a machine that might someday be able to calculate weather forecasts, the mathematician had only imagined predictions that extended a few days into the future, far enough perhaps to bring ships into harbor before a hurricane or prepare a bustling city for a coming blizzard. Richardson would no doubt be amazed to see the state of "numerical processing" two decades into the twenty-first century: machines like the supercomputer "Cheyenne" housed in the Wyoming offices of the National Center for Atmospheric Research, which uses its vast computational power to simulate the behavior of Earth's climate itself. Machines like Cheyenne allow us to simulate time scales that would have seemed preposterous to Richardson: decades, even centuries. The forecasts are fuzzier, of course: you can't ask Cheyenne to tell you whether New Yorkers should dress for rain on July 13, 2087. They can only tell us long-term trends—where new deserts may form, where floods may become more likely, where ice caps may melt—and even those are just probabilities. But that foresight, hazy as it may sometimes seem, is far more accurate than anything Richardson could have imagined just a century ago.

Digital technology is often blamed for the abbreviated attention spans of Snapchat and Twitter, but the fact is that computer simulations have been essential in forcing humans to confront what may be the most complex, long-term decision we have ever faced: what to do about climate change. The near-universal consensus among scientists that global warming poses a meaningful threat has emerged, in large part, thanks to the simulations of supercomputers like Cheyenne. Without the full-spectrum models that those machines are capable of building—tracking everything from planet-scale phenomena like the jet stream all the way down to the molecular properties of carbon dioxide—we would have far less confidence about the potential danger from climate change and the long-term importance of shifting to renewable energy sources. Those simulations now influence millions of decisions all across the planet, from individual choices to buy a hybrid automobile instead of a gas-powered one and community decisions to install solar panels to power public schools all the way up to decisions on the scale of signing the Paris climate accord, truly one of the most global agreements—both in its signatories and its objectives—ever reached in the history of our species.

The fact that we are capable of making these decisions should not be an excuse to rest on our laurels. I am writing these words in the fall of 2017, just a few months after the Trump administration announced that the United States would be withdrawing from the Paris Agreement. It is possible that we will look back at this period twenty or thirty years from now and see this as the beginning of a great unraveling, with more and more citizens dismissing climate change as "fake news," generating increasing paralysis on a governmental level, and undermining efforts to reduce the impact of global warming.

If you polled most Americans, I suspect a majority of them would say that we are getting *worse* at long-term decisions, that we live in a short-attention-span age that keeps us from the long view. A significant number would probably point to the damage we are doing as a species to the environment as the most conspicuous example of our shortsightedness.

It is true that the last few decades have witnessed a number of troubling trends, most of which revolve around that critical attribute of diversity, that have compromised the way we make collective decisions in material ways. In the United States, gerrymandering reduces the ideological diversity behind the decision of who to elect to represent a district in the House of Representatives: members of Congress are increasingly elected by voting blocs that are overwhelmingly Republican or Democratic, far more homogeneous in their political worldviews than most congressional districts would have been at other points of our history. But that trend is not solely attributable to the schemes of politicians trying to ensure reelection. We are also experiencing a demographic "Big Sort," in which our cities and inner-ring suburbs are increasingly populated by Democrats, while Republicans dominate the exurbs and the countryside. So when we come together to make any kind of local decision, we are—politically, at least—assembling teams of decision-makers that are more homogeneous and thus prone to all the failings that homogeneity brings to group decisions.

This is an often underappreciated point in the cultural debates about the importance of diversity. When we look at those images of a Trump cabinet meeting or the Republican House Caucus—all those middle-aged white men in their suits and ties—we tend to frame the lack of diversity in those groups as a problem for egalitarian or representational reasons. And that's a perfectly valid framing. We want a cabinet that "looks like America" because that will get us closer to a world where talented people from all walks of life can find their way into the top echelons of government, and because those different walks of life will inevitably have different interests that will need to be reflected in the way we are governed. But there is another factor that we often ignore when we complain about the lack of diversity at the top of any organization in the private or public sector: *Diverse groups make smarter decisions*. Nowhere is the data on this clearer than in the research on gender and decision-making. If you were trying to assemble a kind of *Springtime for Hitler* anti–dream team, designed to *fail* at complex decisions, you would do well to recruit an all-male roster. So when we see a phalanx of guys signing a bill to block funding to Planned Parenthood, we should not just point out that a woman might have an understanding of Planned Parenthood's value that a man might lack. We should also point out that a group of men is more likely to make the wrong choice about *anything*, not just "women's issues."

But despite those limitations and setbacks, we should remind ourselves that in many other realms, we are attempting to make decisions that involve time horizons and full-spectrum maps that would have been unthinkable to our great-grandparents. No one in 1960 made a decision that contemplated for even a second that decision's impact on atmospheric carbon in 2060. Today, countless people around the globe make decisions that factor in those long-term impacts every single day, from politicians proposing new regulations that include the true cost of carbon in their cost-benefit analysis and corporate executives choosing to run their headquarters on renewable energy sources all the way down to ordinary consumers who choose to buy "green" products at the supermarket.

Think back to Meadow Lake in Queens and those fish struggling to find oxygen beneath the superbloom of blue-green algae. Those fish were, in a sense, stakeholders in the decision process. They were included as a meaningful variable in part because they play an important role in the ecosystem, which ultimately sustains life for human beings, but also because many of us believe they have some intrinsic right to life as a species on this planet, whether they support human needs or not. When the early Manhattanites decided to bury Collect Pond, no one mapped out the impact pathways on the ecology of Lower Manhattan. They just thought they would get rid of an increasingly polluted lake and build some new houses.

Skeptics will argue that, yes, there are some environmental planners out there who are concerned with wetlands wildlife, but if you look at the planet as a whole, we are trashing it at an unprecedented clip. The last two centuries have clearly been the most environmentally destructive of any in human history: for every fish we preserved in Meadow Lake, there are a thousand species we have driven to extinction. Isn’t this clear evidence that we are making worse choices in the modern age?

But the truth is, on a species level, we have been as destructive ecologically as our technology would allow for at least twenty thousand years, maybe longer. No doubt there were some preindustrial communities who factored the “balance of nature” into their collective decisions about what to eat and where to live. But for most of human history, we have been willing to sacrifice just about any natural resource if it aided our short-term needs. Consider the list of mammals driven into extinction during the first few thousand years that humans occupied North America, from roughly 11,000 to 8000 BC: mastodons, jaguars, woolly mammoths, saber-toothed cats, and at least a dozen other species of bears, antelopes, horses, and other animals. For most of our history, our carnage has been reined in far more by our technological limitations than by our intellectual or moral ones. We’ve always churned through everything our tools have allowed us to. We just have better tools now—if “better” is the right word for it—so we can do more damage.

The fish in Meadow Lake, on the other hand, suggest a new kind of deliberation: the decision to preserve a species even if it provides little value to us, in the short term, at least. People have been burying the pond since the Stone Age gave them tools to dig with. But contemplating the impact of nitrogen runoff on an algae bloom and how that bloom might starve the fish of oxygen—that is a new way of thinking.

The fact that some of us continue to debate whether global warming is even happening—let alone what we should do about it—shows us that we're still not experts at this kind of thinking. Yes, it does seem ominous that the United States is currently threatening to withdraw from the Paris climate accord. But we are very early in that particular narrative; the ending is not at all clear. So far, the Paris Agreement story is really the story of two distinct decisions: 198 nations signing the accord itself, and one temperamental leader promising to withdraw in a huff. Approached from the long view, which one looks more impressive? We've had impetuous leaders since the birth of agriculture; truly global accords with real consequences for everyday life are a new concoction.

The fact that we sometimes seem incompetent at these kinds of choices is a sign that we are grading on a reverse curve: we have higher standards now, so it sometimes seems as though we're less deliberative than our ancestors. But the truth is, both the spectrum of our decisions and their time horizons have widened dramatically over the past few centuries. The Aztecs and the Greeks could peer into the future as far as their calendars and their crude astronomy would allow them. They built institutions and structures designed explicitly to last centuries. But they never contemplated decisions that addressed problems that wouldn't arrive for another fifty years. They could see cycles and continuity on the long scale. But they couldn't anticipate emergent problems.

We are better predictors of the future, and our decisions are beginning to reflect that new ability. The problem is that the future is coming at us faster than ever before.

THE LONG VIEW

How long could our time horizons be extended? As individuals, almost all of us will find ourselves contemplating at least a few decisions that by definition extend the length of our lives: who to marry, whether to have children, where to live, what vocation to pursue. As a society we are actively deliberating decisions with time horizons that extend beyond a century, in climate change, automation and artificial intelligence, medicine, and urban planning. Could the horizon recede even farther?

Consider a decision that most of us probably do not, initially, at least, have strong feelings about either way: Should we talk to intelligent life-forms living on other planets? In 2015, a dozen or so science and tech luminaries, including Elon Musk, signed a statement that answered that question with a vehement no: “Intentionally signaling other civilizations in the Milky Way Galaxy,” the statement argued, “raises concerns from all the people of Earth, about both the message and the consequences of contact. A worldwide scientific, political and humanitarian discussion must occur before any message is sent.” They argued, in effect, that an advanced alien civilization might respond to our interstellar greetings with the same graciousness that Cortés showed the Aztecs. The statement was a response to a growing movement led by a multidisciplinary group of astronomers, psychologists, anthropologists, and amateur space enthusiasts that aims to send messages specifically targeting planets in the Milky Way that are likely to support life. Instead of just scanning the skies for signs of intelligent life, the way SETI’s telescopes do, this new approach, sometimes called METI (Messaging Extraterrestrial Intelligence), actively tries to initiate contact. The METI organization, led by former SETI scientist Douglas Vakoch, has planned a series of messages to be broadcast from 2018 onward. And Yuri Milner’s Breakthrough Listen endeavor has also promised to support a “Breakthrough Message” companion project, including an open competition to design the messages to be transmitted to the stars. Think of it as a kind of intergalactic design charrette.

If you believe that the message has a plausible chance of making contact with an alien intelligence, it’s hard not to think of it as one of the most important decisions we will ever make as a species. Are we going to be galactic introverts, huddled behind the door listening for signs of life outside? Or are we going to be extroverted conversation starters? (And if it’s the latter, what should we say?) The decision to send a message into space may not generate a meaningful outcome for a thousand years, or even a hundred thousand years, given the transit times between the correspondents. The first intentional message ever sent—the famous

Arecibo Message sent by Frank Drake in the 1970s—was addressing a cluster of stars fifty thousand light-years away. The laws of physics dictate the minimum time for the result of that decision to become perceptible to us: one hundred thousand years. It is hard to imagine a decision confronting humanity with a longer leash on the future.

The anti-METI movement is predicated on the fact that if we do ever manage to make contact with another intelligent life-form, almost by definition our new pen pals will be far more advanced than we are. (A less advanced civilization would be incapable of detecting our signal, and it would be a staggering coincidence if we happened to make contact with a civilization that was at the same level of technological sophistication as ours.) It is this asymmetry that has convinced so many future-minded thinkers that METI is a bad idea. The human history of exploitation weighs heavily on the imagination of the METI critics. Stephen Hawking, for instance, announced in a 2010 documentary series, “If aliens visit us, the outcome would be much as when Columbus landed in America, which didn’t turn out well for the Native Americans.” Astronomer and sci-fi author David Brin echoes the Hawking critique: “*Every single case* we know of a more technologically advanced culture contacting a less technologically advanced culture resulted at least in pain.”

There is something about the METI decision that forces the mind to stretch beyond its usual limits. Using your own human intelligence, you have to imagine some radically different form of intelligence. You have to imagine time scales where a decision made in 2017 might trigger momentous consequences ten thousand years from now. The sheer magnitude of those consequences challenges our usual measures of cause and effect. If you think METI has a reasonable chance of making contact with another intelligent organism somewhere in the Milky Way, then you have to accept that this small group of astronomers and science-fiction authors and billionaire patrons may, in fact, be wrestling with a decision that could prove to be the most transformative one in the history of human civilization.

All of which takes us back to a much more down-to-earth but no less challenging question: *Who gets to decide?* After many years of debate, the SETI community established an agreed-upon procedure that scientists and government agencies should follow in the event that SETI actually stumbles upon an intelligible signal from space. The protocols specifically ordain that "no response to a signal or other evidence of extraterrestrial intelligence should be sent until appropriate international consultations have taken place." But an equivalent set of guidelines does not yet exist to govern our own interstellar outreach.

The METI debate runs parallel to other existential decisions that we will be confronting in the coming decades, as our technological and scientific powers increase. Should we create superintelligent machines that exceed our own intellectual capabilities by such a wide margin that we cease to understand how their intelligence works? Should we "cure" death, as many Silicon Valley visionaries are proposing? Like METI, these are potentially among the most momentous decisions human beings will ever make, and yet the number of people actively participating in that decision—so far—is minuscule.

One of the most thoughtful participants in the debate over the METI decision, Kathryn Denning, an anthropologist at York University in Toronto, has argued that decisions like METI require a far wider sample of stakeholders: "I think the METI debate may be one of those rare topics where scientific knowledge is highly relevant to the discussion, but its connection to obvious policy is tenuous at best, because in the final analysis, it's all about how much risk the people of Earth are willing to tolerate . . . and why exactly should astronomers, cosmologists, physicists, anthropologists, psychologists, sociologists, biologists, scifi authors, or anyone else (in no particular order) get to decide what those tolerances should be?"

Agreements like SETI protocols—and even the Paris climate accord—should be seen as genuine achievements in the history of human decision-making. But they are closer to norms than to actual legislation. They do not have the force of law behind them. Norms are powerful things. But as we have seen in recent years, norms can also be fragile, easily undermined by disrupters who don't mind offending the mainstream. And they are rarely strong enough to resist the march of technological innovation.

The fragility of norms may be most apparent in decisions that involve extinction-level risk. New technologies (like self-replicating machines) or interventions (like METI) that pose even the slightest risk to our survival as a species require much more global oversight. Creating those regulations would force us to, as Denning suggests, measure risk tolerance on a planetary level. They would require a kind of global Bad Events Table, only instead of calculating the risk magnitude of events that would unfold in a matter of seconds, as the Google algorithm does, the table would measure risk for events that might not emerge for centuries. If we don't build institutions that can measure that risk tolerance, then by default the gamblers will always set the agenda, and the rest of us will have to live with the consequences. This same pattern applies to choices that aren't as much about existential risk as they are about existential *change*. Most Americans and Europeans, when asked, say they would not like to "cure" death; they say they much prefer pursuing longer, more meaningful lives, not immortality. But if immortality is, in fact, within our reach technologically—and there is at least some persuasive evidence to suggest it is—we don't necessarily have the institutions in place that are equipped to stop it. Do we want to have the option to live forever? That is a global, species-level decision if there ever was one.

How would we go about making decisions like this? We do have institutions like the United Nations that gave us a framework for making planetary choices, and for all the limitations of its power, the fact that the UN exists at all is a measure of real progress. If our decision-making prowess improves with the growing diversity of the group making the decision, it's hard to imagine a more farsighted institution than one that represents all the countries of the world. But, of course, the United Nations represents the citizens of those countries through very indirect means. Its decisions are hardly direct expressions of the "will of the people." Would it be possible to conduct something equivalent to a design charrette on the scale of the planet, where stakeholders—not just political appointees—can weigh in with their own priorities and tolerance for risk?

We invented the institution of democracy—in all its many guises—to help us decide, as a society, what our laws should be. Perhaps it is time that we took some of the lessons we have learned from small-group decision-making and applied them to the realm of mass decisions. This is not as unlikely as it sounds. After all, the rise of the Internet has enabled us to reinvent the way we communicate multiple times in my lifetime alone: from email to blogs to Facebook status updates. Why shouldn't we take this opportunity to reinvent our decision-making tools as well?

There is some evidence that Internet crowds can be harnessed to set priorities and suggest options with more acumen than the so-called experts, if the software tools organizing all that collective intelligence (and stupidity) are designed properly. In the month leading up to the 2008 inauguration, the incoming Obama administration opened up a Citizen's Briefing Book on the web, inviting the US population to suggest priorities for the next four years—a small experiment in direct democracy inspired by the Open Government movement that was then on the rise. Ordinary citizens could suggest initiatives and also vote to support other initiatives. In the end, two of the three most popular initiatives urged Obama to radically reform our draconian drug laws and end marijuana prohibition. At the time, the results provoked titters from the media establishment: This is what happens when you open the gates to the Internet crazies; you get a horde of stoners suggesting policy that has zero chance of mainstream support. And yet by the end of Obama's second term, that briefing book turned out to be the first glimmer of an idea whose time had come. Sentencing laws were being rewritten, cannabis was legal in half a dozen states, and a majority of Americans now support full legalization.

In a polarized, nationalistic age, the idea of global oversight on any issue, however existential the threat it poses, may sound naive. And it may well be that technologies have their own inevitability, and we can only rein them in in the short run. Reducing our carbon footprint, by comparison, may prove to be an easier choice than stopping something like METI or immortality research, because there is an increasingly visible path for minimizing climate change that involves adopting even more advanced technology: not retreating back to a preindustrial life, but moving forward into a world of carbon-neutral technology, like solar panels and electric vehicles. In our history, there is not a lot of precedent of human beings voluntarily swearing off a new technological capability—or choosing not to make contact with another society—because of some threat that might not arrive for generations. But maybe it's time we learned how to make that kind of decision.

SUPERINTELLIGENCE

The development of supercomputers like Cheyenne—computers smart enough to map the impact pathways of climate change a hundred years into the future—has endowed us with two kinds of farsightedness: they let us predict future changes in our climate that help us make better decisions about our energy use and our carbon footprint today, and they suggest long-term trends in the development of artificial intelligence, trends that may pose their own existential threat to humans in the coming centuries. The upward trajectory of Moore's law and recent advances in machine learning have convinced many scientists and technologists that we must confront a new global decision: what to do with the potential threat from "superintelligent" machines. If computers reach a level of intelligence where they can outperform humans at nuanced decisions like rendering a verdict in a complicated criminal trial, they will almost certainly have been programmed by evolutionary algorithms, where the code follows a kind of vastly accelerated version of Darwin's natural selection. Humans will program some original base of code, and then the system will experiment with random variations at blistering speed, selecting the variants that improve the intelligence of the machine and mutating that new "species" of code. Run enough cycles and the machine may evolve an intellectual sophistication without any human programmer understanding how the machine got so smart. In recent years, a growing number of scientists and tech-sector leaders—Bill Gates, Elon Musk, Stephen Hawking—have sounded the alarm that a superintelligent AI could pose a potential "existential threat" to humanity.

All of which suggests that we are going to confront a decision as a planet: Are we going to allow superintelligent machines, or not? It's possible that we will "make" the decision in the same way the citizens of New York made the decision to fill Collect Pond, or the way the inventors of the industrial age decided to fill the atmosphere with carbon. In other words, we'll make it in an entirely unstructured, bottom-up way, without any of the long-term deliberation the decision warrants. We'll keep opting for smarter and smarter computers because in the short term, they're better at scheduling meetings and curating workout playlists and driving our cars for us. But those choices won't reflect the potential long-term threat posed by superintelligent machines.

Why would these machines be so dangerous? To understand the threat, you need to shed some of your human biases about the scales of intellectual ability. As AI theorist Eliezer Yudkowsky puts it, we have a “human tendency to think of ‘village idiot’ and ‘Einstein’ as the extreme ends of the intelligence scale, instead of nearly indistinguishable points on the scale of minds-in-general.” From the point of view of, say, a mouse, the village idiot and Einstein are both unfathomably intelligent. We spent the first decades of AI research mostly dreaming about building machines that might function at a village-idiot level of intelligence or perhaps reach the Einsteinian summit. But as the philosopher Nick Bostrom and Yudowsky both argue, there’s no reason to think that the Einsteinian summit is some sort of fundamental upper limit. “Far from being the smartest possible biological species,” Bostrom writes, “we are probably better thought of as the stupidest possible biological species capable of starting a technological civilization—a niche we filled because we got there first, not because we are in any sense optimally adapted to it.” Powered by recursive, self-learning algorithms, the first true AI might well march right past Mount Einstein and ascend to some higher plateau well beyond our imagining.

The danger perceived by people like Bostrom or Hawking does not look exactly like the standard science-fiction version. First, it is not at all necessary that the AI become conscious (or “self-aware,” as the original *Terminator* put it). A superintelligent AI might develop some kind of alternative consciousness, likely completely different from ours. But it also might remain a vast assemblage of insentient calculations, capable of expression and action and long-term planning, but lacking a sense of self. Secondly, the AI need not suddenly turn evil or vengeful or ambitious (or any other anthropomorphic emotion) to destroy human civilization. Bostrom, for instance, spends almost no time in his influential book *Superintelligence* imagining machines becoming evil overlords; instead, he worries about small miscommunications in defining the AI’s goals or motivations that could lead to global or even cosmic transformations. Consider programming an AI with as seemingly an innocuous goal as you could imagine: Bentham’s “greatest happiness for the greatest number.” You set that as the overarching value and let the machine decide the best approach to making it a reality. Maximizing human happiness would seem to be a perfectly laudable objective, but the AI might well come up with a scenario that, while technically achieving the objective, would be immediately abhorrent to humans: perhaps the AI distributes nanobots into every human brain on the planet, permanently stimulating the pleasure centers of the brain and turning us all into grinning zombies. The threat is not that when asked to decide the best strategy for combating some environmental crisis, the AI will actively disobey us and instead hack into the Department of Defense network and detonate its entire nuclear arsenal because it has evolved some inherent evilness or desire for conquest. The threat is that we will ask it to find the optimal solution for an environmental crisis, and it will decide to eliminate the main cause of the crisis—human beings—because we haven't framed the objective clearly enough.

Much of the debate over superintelligent AI is devoted to thinking through what is sometimes called the "containment problem," brilliantly explored in Alex Garland's film *Ex Machina*: how to keep the genie of AI inside the bottle, while still tapping into its powers. Could humans evolve an AI that was truly superintelligent, but at the same time keep it safely bounded so that a runaway instruction doesn't trigger a global catastrophe? In Bostrom's convincing presentation, the problem is much harder than it might first appear, in large part because the humans would be trying to outthink an intelligence that is orders of magnitude more advanced than their own. Containing the AI will be like a mouse scheming to influence human technological advancement to prevent the invention of mousetraps.

In a way, we are at a point in the conversation about superintelligence equivalent to where the global warming debate was in the late 1980s: a small group of scientists and researchers and public intellectuals extrapolating out from current trends and predicting a major crisis looming several generations down the line. According to a survey conducted by Bostrom, most of the AI research community believes superhuman-level AI is still at least fifty years away.

That multigenerational time scale may be the most encouraging element in a debate filled with doomsday scenarios. Climate advocates often complain about the sluggish pace of political and corporate reform given the magnitude of the global warming threat. But we should remind ourselves that with climate change, we are trying to make a series of decisions that are arguably without precedent in human history: deciding which regulatory and technological interventions to put in place to prevent a threat that may not have a severe impact on most humans for several decades, if not longer. For all the biases and intuitive leaps of System 1, one of the hallmarks of human intelligence is the long-term decision-making of System 2: our ability to make short-term sacrifices in the service of more distant goals, the planning and forward thinking of *Homo prospectus*. While we are not flawless at it by any means, we are better at that kind of thinking than any other species on the planet. But we have never used those decision-making skills to wrestle with a problem that doesn't exist yet, a problem we anticipate arising in the distant future based on our examination of current trends.

To be clear, humans have made decisions to engineer many ingenious projects with the explicit aim of ensuring that they last for centuries: pyramids, dynasties, monuments, democracies. Some infrastructure decisions—like the dike system of the Netherlands or Japanese building codes designed to protect against tsunamis—anticipate threats that might not happen for a century or more, though those threats are not genuinely new ones: those cultures know to be concerned about floods and tsunamis because they have experienced them in the past. Some decisions that we have made, like the decision to adopt democratic governance, have been explicitly designed to solve as-of-yet-undiscovered problems by engineering resilience and flexibility into their codes and conventions. But mostly those exercises in long-term planning have been all about preserving the current order, not making a preemptive choice to protect us against threats that might erupt three generations later. In a way, the closest analogues to the current interventions on climate (and the growing AI discussion) are eschatological: in religious traditions that encourage us to make present-day decisions based on an anticipated Judgment Day that may not arrive for decades, or millennia.

With superintelligence, as with climate change, we are trying something new as a species. We are actively thinking about the choices we are making *now* in order to achieve a better outcome fifty years from now. But superintelligence is an even more ambitious undertaking, because the problem we are anticipating is qualitatively different from today's reality. Climate change forces us to imagine a world that is a few degrees hotter than our current situation, with longer droughts, more intense storms, and so on. We talk about global warming "destroying the planet," but that language is hyperbole: the planet will be fine even if we do nothing to combat global warming. Even in the worst-case scenario, *Homo sapiens* as a species would survive a five-degree increase in surface temperatures—though not without immense suffering and mortality. A truly superintelligent machine—capable, for example, of building self-replicating nano-machines that devour all carbon-based life—could plausibly pose an extinction-level threat to us. But there is nothing in our current landscape or our history that resembles this kind of threat. We have to imagine it.

Interestingly, one of the key tools we have had in training our minds to make this momentous choice has been storytelling—science fiction, to be precise, which turns out to play a role in some of our mass decisions equivalent to the role scenario planning plays in our group decisions. "This kind of exercise is generally new," the writer and futurist Kevin Kelly suggests, "because we all now accept that the world of our grandchildren will be markedly different than our world—which was not true before. I believe this is the function of science fiction. To parse, debate, rehearse, question, and prepare us for the future of new. For at least a century, science fiction has served to anticipate the future . . . In the past there have been many laws prohibiting new inventions as they appeared. But I am unaware of any that prohibited inventions before they appeared. I read this as a cultural shift from science fiction as entertainment to science fiction as infrastructure—a necessary method of anticipation." Science-fiction narratives have been ruminating on the pitfalls of artificial intelligence for at least a century—from the "global brain" of H. G. Wells to HAL 9000 all the way to *Ex Machina*—but only in the last few years has the problem entered into real-world conversation and debate. The novels primed us to see the problem more clearly, helped us peer around the limits of our technology-bounded rationality. No doubt the superintelligent machines will climb their way past human intelligence by running ensemble simulations at unimaginable speeds, but if we manage to keep them from destroying life as we know it, it will be, in part, because the much slower simulations of novels helped us understand the threat more clearly.

Given the accelerating rates of change of modern society, the current debate about AI and its potential threats is a bit like a group of inventors and scientists gathering together in the early 1800s and saying, "This industrial revolution is certainly going to make us much more productive, and in the long run raise standards of living, but we also appear to be pumping a lot of carbon into the atmosphere, which will likely come back to haunt us in a couple of centuries, so we should think about how to prevent that problem.” But, of course, that conversation didn’t happen, because we didn’t have the tools to measure the carbon in the air, or computer simulations to help us predict how that carbon would influence temperatures globally, or a history of battling other industrial pollutants, or government and academic institutions that monitor climate and ecosystem change, or sci-fi novels that imagined a scenario where new technologies somehow altered global weather patterns. We were smart enough to invent coal-powered engines, but not yet smart enough to predict their ultimate impact on the environment.

The AI debate is yet another reminder of how much progress we have made in our ability to make farsighted decisions. All those tools and sensors and narratives that have enabled us to identify the threat of climate change or imagine an AI apocalypse constitute, en masse, their own kind of superintelligence.

“We are as gods,” Stewart Brand famously wrote a half century ago. “We might as well get good at it.” We have indeed developed godlike powers over our planet’s atmosphere in just under three hundred years of carbon-based industry. Are we good at it yet, though? Probably not. But we’re quick learners. And we’re certainly taking on global decisions with time horizons that our immediate ancestors would have found astonishing. The fact that challenges are still emerging to long-term global decisions like the Paris Agreement is inevitable: it’s hard enough to project forward fifty years as an individual, much less as a society. But just the existence of these debates—AI, climate change, METI—make it clear that we are beginning to explore a new kind of farsightedness. With AI, all the projections of future threats may well turn out to be false alarms, either because true AI turns out to be far more difficult to achieve, or because we discover new techniques that minimize the danger before the machines march on past Mount Einstein. But if artificial superintelligence does turn out to pose an existential threat, our best defense will likely come out of our own new powers of *human* superintelligence: mapping, predicting, simulating, taking the long view.

THE DRAKE EQUATION

Superintelligence, climate change, and METI share another property beyond their extended time horizons. They are all decisions that cannot be properly appraised without the consultation of a wide range of intellectual disciplines. Climate science alone is a hybrid of multiple fields: molecular chemistry, atmospheric science, fluid dynamics, thermodynamics, hydrology, computer science, ecology, and many more. Defining the problem of climate change didn’t just require the digital simulations of Cheyenne; it also required a truly heroic collaboration between disciplines. But deciding what to do about climate change requires a whole other set of fields as well: political science, economics, industrial history, and behavioral psychology, for instance. The problem of superintelligence draws on expertise in artificial intelligence, evolution, and software design, but it also has been profoundly illuminated by philosophical inquiries and the imagined futures of science fiction. Some amount of intellectual diversity is required in any full-spectrum decision, of course; even the most intimate choice, as we will see in the next chapter, draws on multiple bands of experience to settle on an optimal path. But these mass decisions—the ones that may well involve existential risk to us as a species—require an even wider slice of the spectrum.

A little more than a decade before he transmitted his famous Arecibo Message—the one that cannot, by definition, receive a reply for another hundred thousand years—Frank Drake sketched out one of the great equations in modern scientific history, as a way of framing the decision of whether to seek contact with lifeforms on other planets. If we start scanning the cosmos for signs of intelligent life, Drake asked, how likely are we to actually detect something? The equation didn’t generate a clear answer; it was more of an attempt to build a full-spectrum map of all the relevant variables. In mathematical form, the Drake equation looks like this:

$$N = R_* \times f_p \times n_e \times f_l \times f_i \times f_c \times L$$

N represents the number of extant, communicative civilizations in the Milky Way. The initial variable R_* corresponds to the rate of star formation in the galaxy, effectively giving you the total number of potential suns that could support life. The remaining variables then serve as a kind of nested sequence of filters: Given the number of stars in the Milky Way, what fraction of those have planets, and how many of those have an environment that can support life? On those potentially hospitable planets, how often does life itself actually emerge, and what fraction of that life evolves into intelligent life, and what fraction of that life eventually leads to a civilization's transmitting detectable signals into space? At the end of his equation, Drake placed the crucial variable *L*, which is the average length of time during which those civilizations emit those signals.

I know of no other equation that so elegantly conjoins so many different intellectual disciplines in a single framework. As you move from left to right in the equation, you shift from astrophysics, to the biochemistry of life, to evolutionary theory, to cognitive science, all the way to theories of technological development. Your guess about each value in the Drake Equation winds up revealing a whole worldview. Perhaps you think life is rare, but when it does emerge, intelligent life usually follows; or perhaps you think microbial life is ubiquitous throughout the cosmos, but more complex organisms almost never form. The equation is notoriously vulnerable to very different outcomes, depending on the numbers you assign to each variable.

The most provocative value is the last one: *L*, the average life span of a signal-transmitting civilization. You don't have to be a Pollyanna to defend a relatively high *L* value. You just have to believe it's possible for civilizations to become fundamentally self-sustaining and survive for millions of years. Even if one in a thousand intelligent life-forms in space generates a million-year civilization, the value of *L* increases meaningfully. But if your *L* value is low, that implies a further question: What is keeping it low? Do technological civilizations keep flickering on and off in the Milky Way, like so many fireflies in space? Do they run out of resources? Do they blow themselves up?

Since Drake first sketched out the equation in 1961, two fundamental developments have reshaped our understanding of the problem. First, the product of the first three values in the equation (representing our best guess at the number of stars with habitable planets) has increased by several orders of magnitude. And second, we have been listening for signals for decades and heard nothing. If the habitable planet value keeps getting bigger and bigger without any sign of intelligent life in our scans, the question becomes: Which of the other variables are the filters? Perhaps life itself is astonishingly rare, even on habitable planets. From our perspective, as human beings living in the first decades of the third millennium, wondering whether we are flirting with existential risks through our technological hubris, we want the emergence of intelligent life to be astonishingly rare; if the opposite is true, and intelligent life is abundant in the Milky Way, then *L* values might be low, perhaps measured in centuries and not even millennia. In that case, the adoption of a technologically advanced lifestyle might be effectively simultaneous with extinction. First you invent radio, then you invent technologies capable of destroying all life on your planet and shortly thereafter you push the button and your civilization goes dark.

Perhaps this is the ironic fate of any species that achieves the farsightedness of *Homo prospectus*. Perhaps every time a species on some Earth-like planet evolves a form of intelligence smart enough to imagine alternate futures, smart enough to turn those imaginative acts into reality, that cognitive leap forward sets off a chain reaction of technological escalation that ultimately deprives that species of its actual future. The early silence that has greeted our SETI probes so far suggests that this is at the very least a possibility. But perhaps that escalation is an arms race that is not doomed to end in apocalypse. Maybe the *L* values are high, and the universe is teeming with intelligent life that made it through the eye of the needle of industrialization without catastrophe. Maybe it’s possible to invent ways of making farsighted choices as a society faster than we invent new ways of destroying ourselves. Certainly it’s essential for us to try. If those superintelligent machines do manage to assist human civilization, and not accidentally trigger the mass extinction that Bostrom and Hawking fear, it will be because those machines learned how to make decisions that assessed the full spectrum of variables and consequences, that ran ensemble simulations that allowed them to tease out all the unanticipated consequences and discover new options. Perhaps the machines will evolve that farsightedness on their own, through some kind of self-learning algorithm. But wouldn’t it be better if we were wise enough by then to give them a head start?