

PREVAILING OVER OUR DYSTOPIAS

Joel Garreau*

Lord Martin Rees is the “astronomer royal” of the United Kingdom¹—sort of a poet laureate to the stars. From this lofty vantage point, Rees has produced the most far-reaching imaginable catalog of dystopian outcomes for the human race and the universe. He can even see that “[e]xperiments that crash atoms together with immense force could start a chain reaction that erodes everything on Earth; the experiments could even tear the fabric of space itself, an ultimate ‘Doomsday’ catastrophe whose fallout spreads at the speed of light to engulf the entire universe.”²

In his book *Our Final Hour: A Scientist’s Warning: How Terror, Error and Environmental Disaster Threaten Humankind’s Future in This Century—On Earth and Beyond*, Rees vociferously makes the case that “technical advances will in themselves render society more vulnerable to disruption.”³ He quotes the odds of our species surviving to the end of the twenty-first century as “no better than fifty-fifty.”⁴

Chapter by chapter, Rees catalogs everything that could go colossally wrong in this century. It is hard to believe he’s missed anything. Climate change, asteroid impact, flesh-eating viruses assembled and released by madmen, genetic fiddling run amok, nanobots run amok, artificial intelligence (AI) run amok—it is all in there. So are quite a few exotic catastrophes that could theoretically result from experiments in Rees’ specialty, physics—such as the universe tearing mentioned above.

As expert and devastating as is his recitation, Rees has difficulty with the issue of what to do about these doomsday scenarios. He does not like risk taking

*Joel Garreau, the author of *Radical Evolution: The Promise and Peril of Enhancing Our Minds, Our Bodies—And What It Means to Be Human*, is Professor of Law, Culture and Values at Arizona State University’s Sandra Day O’Connor College of Law. He would like to acknowledge his colleagues at Stratfor Global Intelligence for giving him an opportunity to refine this argument with them.

1. *Martin Rees: Biography*, ROYAL SOC’Y, <https://royalsociety.org/people/martin-rees-12156/> (last visited July 3, 2016).

2. MARTIN REES, *OUR FINAL HOUR: A SCIENTIST’S WARNING; HOW TERROR, ERROR, AND ENVIRONMENTAL DISASTER THREATEN HUMANKIND’S FUTURE IN THIS CENTURY—ON EARTH AND BEYOND* 1–2 (2003).

3. *Id.* at 21.

4. *Id.* at 8.

at all. Take our world today. If it were up to him, we would not be living in it. Given the risks of a nuclear exchange during the Cold War, Rees flatly states he would “rather be red than dead,”⁵ in the words of the old slogan. “I personally would not have chosen to risk a one in six chance of a disaster that would have killed hundreds of millions and shattered the physical fabric of all of our cities, even if the alternative was a certainty of a Soviet takeover of Western Europe,” he says.⁶

That’s the problem with dystopianism. Once you start to immerse yourself, it’s easy to drown in it. Or at least never think past it. As we are presently seeing in parts of Silicon Valley.

For almost a decade, the dominant Silicon Valley prediction has been Singularity Utopianism. In this story about the future, the godlike powers afforded by the genetics, robotics, AI, and nano revolutions rapidly cure stupidity, ignorance, pain, suffering, and even death. We merge with our machines and thus transcend. This outcome is inevitable, according to this prediction, because technology is on its ever-increasing march, and it matters little what we try to do about it. Call this the Heaven Prediction.

Recently, however, the fashionable prediction among the techno elite has flipped to radical dystopianism. The augury of the moment is the imminent arrival of satanic AI that will be the last invention humans ever make, or will be allowed to make. The word *doom* is used liberally. In this reading of the tea leaves, technology is in control and there is frighteningly little we can do about it. Call this the Hell Prediction.

What is up with this astonishing swing? And are these really the only two doors for humanity to pass through?

Let’s be clear: There is nothing wrong with these Heaven and Hell stories as *scenarios*. They are perfectly credible and legitimate possible futures logically based on existing facts. What’s remarkable, though, is that many of the advocates of these futures present them as stone-cold *predictions*. They see no alternative.

In truth, techies’ very deep “super-brain” worries about the accelerating and astonishing powers of AI go back years. But in 2015 came the explosive announcement from tech luminary Elon Musk, renowned physicist Stephen Hawking, and many creators of AI.⁷ They warned that the “intelligence explosion” could sink the human race. Such legends as Apple cofounder Steve Wozniak and Microsoft cofounder Bill Gates soon joined in sharing their concerns.

This is highly reminiscent of the moment back in April 2000, when Bill Joy, cofounder of Sun Microsystems and sometimes called “the Edison of the

5. *Id.* at 28.

6. *Id.*

7. *An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence*, FUTURE LIFE INST., <http://futureoflife.org/ai-open-letter/> (last visited Sept. 27, 2016). See also Stuart Russell et al., *Research Priorities for Robust and Beneficial Artificial Intelligence*, AI MAG., (Winter 2015), http://futureoflife.org/data/documents/research_priorities.pdf (noting the term *intelligent explosion*).

Internet,”⁸ presented his manifesto “Why the Future Doesn’t Need Us.”⁹ It appeared in *Wired*—the house organ of the digerati—and was subtitled “Our most powerful 21st-century technologies—robotics, genetic engineering, and nanotech—are threatening to make humans an endangered species.”¹⁰ It included the AI apocalypse and more. Joy explicitly intended it as a wake-up call comparable in magnitude to that of Albert Einstein advising Franklin D. Roosevelt of the possibility of an atomic bomb. As do Musk, Hawking, and company with their warnings today. And they are not kidding, and they are not wrong, and they are doing the species a favor.

But the surprising thing at this moment—well, maybe it isn’t so surprising—is how often the techies can’t think past their transistors when it comes to the impact of their creations on culture, values, society, and the future of the human race.

Joy set the pattern that others continue to follow. First they pay due attention to Ray Kurzweil. Kurzweil is the polymath author of *The Singularity Is Near: When Humans Transcend Biology* and similar works. He is now a director of engineering at Google, heading up a team that develops machine intelligence. He is also cheerleader-in-chief for the Heaven Prediction at Singularity University, which he cofounded. The University’s stated aim is to “educate, inspire and empower leaders to apply exponential technologies to address humanity’s grand challenges.”¹¹ Musk has served there as a featured presenter.

The next step, however, is for it to occur to the techies, “Hey, wait a minute. This could go exactly the opposite way.” This is the moment—when finally they realize the Heaven Prediction is not bulletproof—that they switch to the other simplistic prediction because they can see no logical alternative. Then they turn against their own creations.

The problem here is that both Heaven and Hell are technodeterministic stories. They are mirror images. Both assume that the core driver of change is how many transistors you can hook up, how fast. They then take the nice smooth curve of Moore’s Law, map it onto the future of the human race—up or down—and *voilà*, they have a prediction. Technology drives history, in this view, leaving little or no room for human agency, and the outcome is inevitable.

We have seen this error before. In the 1950s, no one would have given you a plugged nickel for the scenario we live in today, in which no one has popped a nuclear weapon in anger for 70 years. Of course, that abstinence could change in the next 20 minutes. But we humans for three generations have figured out how to avoid this existential peril—and prevail.

The Hell Prediction folk—like Joy—ascribe that to “luck.” Whenever the species dodges a bullet, they call it sheer blind fortune.

8. Brent Schlender, *The Edison of the Internet Sun Microsystems Mastermind Bill Joy Is on a 20-Year Streak of Innovation That Has Laid the Groundwork for a New Technological Era*, FORTUNE (Feb. 15, 1999), http://archive.fortune.com/magazines/fortune/fortune_archive/1999/02/15/254898/index.htm.

9. Bill Joy, *Why the Future Doesn’t Need Us*, WIRED (Apr. 1, 2000, 12:00 PM), <http://www.wired.com/2000/04/joy-2/>.

10. *Id.*

11. SINGULARITY UNIVERSITY, <http://singularityu.org> (last visited July 4, 2016).

Maybe. But when a species manages to create its own luck for millennia, you have to start wondering how.

Enter the Prevail Scenario. This is the third story for how our futures might go. It is far more faithful to history as we have known it. Prevail is not some middle ground between Heaven and Hell. It is way off in its own territory. Its fundamental assumption is that what matters is not how many transistors you can hook up—a la Moore's Law. It's how many ornery, cussed, imaginative, surprising *humans* you can hook up.

Unquestionably, if we're waiting for House Judiciary or some learned university center to resolve our predicaments at their usual pace, that's a problem. If, at the same time, game-changing challenges to the future of the human race are increasing on a curve, the gap just keeps getting wider.

Suppose, however, that our bottom-up, flock-like human responses to these challenges are also rapidly increasing on a second curve. Then we have a shot.

There's reason for guarded optimism about the existence and efficacy of that second curve. If you look out at the future from 1200 AD, you see marauding hordes and plague, and you say, "Okay, this experiment is over." But then circa 1450 you get movable type and the printing press. All of a sudden you've got a brand-new way for humans to store, share, and distribute their ideas. The results are amazing. First you get the Renaissance. And then the Enlightenment, which yields that massively parallel processing called democracy. And science itself. And you find yourself in our world today, in which 1200 is ancient history in every sense.

These discoveries and innovations were beyond the imagination of any one king or country. These achievements were not top-down. They were bottom-up. Frequently in defiance of power, notably the church. Our world was created by people who came together, collectively, to do the best they could against dire odds. And sometimes hitting transcendence. If you want to call this "heroic muddling-through," I won't argue. Our literature is full of Prevail stories—from *Exodus* in the Bible to the British "nation of shopkeepers" prevailing against the Third Reich. From *Huckleberry Finn* to *Casablanca*. From *Star Wars* to *Lord of the Rings* to *Harry Potter*.

Today you can see it in our headlines. On 9/11, the fourth airplane—Flight 93—never made it to its target. Why? Because the Air Force was so smart? No. Because the White House was so smart? *Hell, no*. It's because a small group of people on board that aircraft—empowered by their airphone technology—figured out, diagnosed, and cured their society's ills in a little under an hour. Was it an ideal solution? No. They all died. But they prevailed.

So how would you know if the Prevail Scenario was the future actually coming into being? Are we seeing an exponential increase in the quantity, quality, variety, and complexity of ways that humans are finding to connect? Are we seeing novel and interesting group behavior as a result—like flocks doing amazing and surprising things?

Well, how about eBay? That's not just the world's biggest flea market. That's over 100 million people worldwide achieving complexity without leaders for a long time (by Internet standards). Wikipedia amplifies our minds. I have

no idea what Twitter is good for, but if it flips out every tyrant in the Middle East, I'm interested.

So here's the question for those facing our dystopias: We know that innovation centers like the Defense Advanced Research Projects Agency (DARPA) can and do accelerate the first curve of technological change. Can we, reading this piece, become the DARPA of the second curve? Can we accelerate our species' coevolution—to our ends?

The central question of this coevolution is not what the computer will become. It's what kind of people we are becoming.

Can human understanding about human understanding increase?

Can we learn what actually makes teams work?

Do we have a moral obligation to use enhancement technology to make ourselves beings who are more compassionate, moral, and wise?

Is that our only chance for survival? As the scenarist Arie de Geus says: "The . . . ability to learn faster (and possibly better) than the competition becomes its most sustainable competitive advantage."¹²

The stakes could not be higher.

We cannot detect any other intelligent life in the universe. It has occurred to me to wonder whether every intelligent species gets to the point where it takes control of its own evolution.

Maybe this is the final exam.

Maybe everybody else flunked.

Let us not flunk.

12. ARIE DE GEUS, *THE LIVING COMPANY* 157 (1997).