Facing Disaster: The Great Challenges Framework

Phil Torres

Published in special issue of *Foresight* on existential risks. Updated Nov. 6, 2018.

Abstract: This paper provides a detailed survey of the greatest dangers facing humanity this century. It argues that there are three broad classes of risks—the "Great Challenges"—that deserve our immediate attention, namely, environmental degradation, which includes climate change and global biodiversity loss; the distribution of unprecedented destructive capabilities across society by dual-use emerging technologies; and value-misaligned algorithms that exceed human-level intelligence in every cognitive domain. After examining each of these challenges, the paper then outlines a handful of additional issues that are relevant to understanding our existential predicament and could complicate attempts to overcome the Great Challenges. The central aim of this paper is to provide an authoritative resource, insofar as this is possible in an academic journal, for scholars who are working on or interested in existential risks. In my view, this is precisely the sort of big-picture analysis that humanity needs more of if we wish to navigate the obstacle course of existential dangers before us.

This is the first century in the world's history when the biggest threat is from humanity.—Lord Martin Rees

1. Introduction

The present paper offers a comprehensive overview of our rapidly evolving existential predicament. In doing this, it outlines what I call the "Great Challenges framework," which aims to provide a useful mechanism for *prioritizing* the various threats to humanity in the first half of the twenty-first century. To qualify as a Great Challenge, a phenomenon must satisfy three criteria: (i) *significance*: it must have existential risk implications; (ii) *urgency*: it must require immediate attention if humanity wishes to obviate it; and (iii) *ineluctability*: it must be more or less unavoidable given civilization's current developmental trajectory. Put differently, the semantic extension of "Great Challenges" includes all and only those problems that humanity cannot ignore, are time-sensitive, and would produce severe and irreversible consequences if left unsolved. For the sake of clarity, we can define an existential risk here as "any future event that permanently prevents us from exploiting a large portion of our cosmic endowment of negentropy to create astronomical amounts of those things that we find valuable" (see Torres and Beard, forthcoming).[1] This reconstructs the various definitions articulated by Bostrom in his foundational work on the topic (e.g., Bostrom 2002, 2013).[2] The most obvious way of satisfying these definiens is for humanity to go extinct, but there are other events, e.g., civilizational collapse and technological stagnation, that could also irreparably compromise our long-term prospects for realizing astronomical value in the universe (see Bostrom 2013).[3]

Why should one care about the long-term survival and flourishing of (post)humanity? The answer to this question goes (far) beyond the scope of this paper, but suffice it to say that many value systems would view the instantiation of an existential risk as tragic. For example, total utilitarianism prescribes the creation of as many happy people in the universe as possible, and since failing to exploit our cosmic endowment would hinder this aim, an existential catastrophe would be profoundly bad (see, e.g., Parfit 1984; Bostrom 2013). In contrast, Johann Frick (2017) argues that humanity has non-instrumental or "final" value and that "when what is finally valuable is a form of life or a species, what we ought to care

---

[1] This gestures at what I have elsewhere termed the "astronomical value thesis," which states that the value of the far future could be astronomically great (Torres 2017a).

[2] Although see Torres and Beard, forthcoming, for some criticisms of Bostrom's lexicographic and typological definitions.

[3] Note that the Great Challenges approach contrasts with the *etiological approach*, which categorizes existential risks according to their proximate causes, as well as the *outcome approach*, which categorizes them according to their consequences. Bostrom (2013) provides an example of the latter whereas Bostrom and Milan Ćirković (2008) and the present author provide a rough outline of risk outcomes organized by their causes (from criticisms of this approach, see Häggström 2016).

about, we might say, is the ongoing instantiation of the universal." It follows that our extinction sooner rather than later "would be very bad, indeed one of the worst things that could possibly happen" (Frick 2017). Finally, Samuel Scheffler (2016, 2018) offers yet another perspective according to which much of what gives current lives value—i.e., makes them "value-laden"—is predicated on an assumption that humanity will survive long into the future. Thus, our lives would become largely meaningless, he claims, if we were to discover that humanity is soon to perish. The point is that there are multiple convergent arguments for why succumbing to an existential catastrophe would be extremely bad (see Torres and Beard, forthcoming).[4] This implies that mitigating existential risks through, as it were, "broad" or "targeted" strategies "should be a dominant consideration whenever we act out of an impersonal concern for humankind as a whole" (Bostrom 2013; see also Beckstead 2013). Bostrom dubs this the "maxipok" rule of thumb, and it is the motivational decision-theoretic component of the Great Challenges framework. That is to say: The maxipok rule instructs us to reduce the overall probability of existential risk, but it doesn't tell us which particular existential risk scenarios we should focus on to most effectively achieve this end; an answer to this prioritization question is what the Great Challenges framework aims to provide.

Before examining the three Great Challenges in some detail, let's begin with a brief survey of (what I called above) "our rapidly evolving existential predicament." First, the good news: Humanity has made significant progress in multiple epistemic, moral, technological, medical, and so on, domains over time. For example, our scientific models of the universe have never been so complete and future anticipated technologies promise to eliminate virtually all human diseases and perhaps reverse the process of aging. Even more, studies reveal an appreciable decline in the prevalence of nearly every form of violence across history, including assaults, rapes, murders, genocides, and wars. Contemporary people stand at the vanguard of the Long Peace, during which no two world superpowers have gone to war, and we are riding the wave of what Steven Pinker (2011) calls the "New Peace," which denotes the period since the end of the Cold War during which "organized conflicts of all kinds—civil wars, genocides, repression by autocratic governments, and terrorist attacks—have declined throughout the world." And the postwar "Rights Revolutions" have ameliorated the plights of "ethnic minorities, women, children, homosexuals, and animals" (Pinker 2011, xxiv). According to Pinker's "escalator hypothesis," the driving force behind these trends in recent times has been (a) the Flynn effect, yielding what he calls the "*moral* Flynn effect," and (b) the propagation of Enlightenment values like reason, science, and humanism (Pinker 2018; although see Torres 2018 for criticisms). From this perspective—sometimes called "New Optimism"—not only has the world improved in numerous important respects, but there appears to be some justification for sanguinity about our collective future in the cosmos.

Yet this cluster of encouraging trends is only half of the diachronic picture of humanity, so to speak. The human condition is indeed better overall today than in the past, but the contemporary world also contains far more *risk potential* than any previous moment in anthropological history.[5] By "risk potential," I mean a rough measure of the extent to which things could go wrong in a global-transgenerational sense; for example, the more existential risk scenarios there are, the greater the risk potential. This leads us to the following two interrelated trends:

First, the *total number* of global-scale catastrophe scenarios has significantly risen since 1945 (Torres 2017a). Prior to the inauguration of the Atomic Age, the only risks to human survival stemmed from natural phenomena like asteroids, comets, supervolcanoes, gamma-ray bursts, solar flares, cosmic rays, and pandemics. Today the list of existing and emerging threats includes a growing constellation of anthropogenic risks like climate change, global biodiversity loss, species extinctions, nuclear conflict, designer pathogens, atomically-precise 3D printers, autonomous nanobots, stratospheric geoengineering,

---

[4] Although see the end of section 3 for some complications arising from the concepts of an "s-risk" and "hyperexistential risk."

[5] Arguably the two most existentially dangerous episodes of our species' life so far were the Toba catastrophe and the Cuban Missile Crisis. Yet the risk potential of the world during the Toba catastrophe—if indeed it nearly wiped out our species as some scientists claim—was still far below the risk potential of today. See the probability estimates of a natural versus anthropogenic existential catastrophe below (see Torres, forthcoming).

physics disasters, and machine superintelligence, to name just a few.[6] And let's not forget that super-powerful future artifacts could introduce entirely new types of risks to human survival and prosperity; these artifacts may be, from our current vantage point, not merely unimagined but *unimaginable*, perhaps requiring a different kind of mind to comprehend. Indeed, I have elsewhere argued that "unknown unknowns"—or, more playfully, "*monsters*," of which there are three types[7]—could constitute the greatest long-term threat to humanity (Torres 2016).[8] Monsters could take the form of novel inventions, cosmic dangers presently hidden from sight, and unintended consequences. If the overall level of risk grows in proportion to the exponential development of new dual-use technologies, then one might even consider talking about an "existential risk singularity," in Ray Kurzweil's (2005) sense of "Singularity" (see Verdoux 2009).

Second, the *overall probability* of a global-scale catastrophe appears to be unprecedentedly high on anthropological timescales.[9] Consider that the probability of annihilation per century from what I call our "*cosmic risk background*." That is to say, the cluster of risks from natural phenomena is almost certainly less than 1 percent; Toby Ord (2015) argues that it may be *much less* than 1 percent, perhaps circa a 1/300 chance per century. By contrast, Sir Nicholas Stern assumes a 0.01 percent chance of extinction each year in his influential Stern Report (2006), which yields a 9.5 percent chance per century, although this number was chosen as a modeling assumption for the purposes of discussing discount rates. The point is that even on high estimates of natural existential risk per century, the overall probability is relatively low. However, when one seriously considers the ballooning swarm of anthropogenic doomsday scenarios, the overall probability appears unsettlingly high. Consider the following figures from scholars of global risk:

> (i) John Leslie estimates that the probability of annihilation in the next 500 years is 30 percent (Leslie 1996).[10]
> (ii) Nick Bostrom writes that his "subjective opinion is that setting this probability [of an existential risk] lower than 25 percent would be misguided, and the best estimate may be considerably higher" (Bostrom 2002).[11] Elsewhere Bostrom conjectures that the "probability that humankind will fail to survive the 21st century" is "not less than 20%" (Bostrom 2005), and in 2014, he *seems* to have assigned a 17 percent probability to the proposition that "humanity goes extinct in the next 100 years" (see Sandberg 2014).
> (iii) Lord Martin Rees puts the likelihood of civilizational collapse before 2100 at 50 percent (Rees 2003).
> (iv) An informal survey of experts conducted by the Future of Humanity Institute (FHI) yielded a median probability of extinction this century at 19 percent (Sandberg and Bostrom 2008).
> (v) Willard Wells uses a mathematical "survival formula" to calculate that, as of 2009, the risk of extinction is almost 4 percent per decade and the risk of civilizational collapse is roughly 10 percent per decade (Wells 2009).
> (vi) Toby Ord estimates that, given the future development of "radical new technology," humanity has a 1/6 chance of going extinct this century (see Wiblin 2017).

---

[6] The possibility that we live in a computer simulation that gets shut down constitutes another interesting possibility: if we are sims, then this particular risk has, as a matter of fact, been with us all along, although we have only recently identified it as a doomsday scenario. Yet the act of identifying it could potentially increase the probability that our simulation gets shut down—or so some scholars have speculated. See Bostrom 2003a.

[7] That is, individual-relative, science-relative, and mind-relative; see Torres 2016.

[8] The term "monster" thus joins a colorful, growing list of similar words, e.g., "black swans," "grey rhinos," "black elephants," and so on.

[9] In an informal discussion about probability and existential risk, Matthijs Maas suggested the neologism "microdoom" or, plural, "microdooms" to indicate something like "a one-in-a-million chance of an existential catastrophe." Colleagues and I plan on elaborating this idea in a future paper.

[10] Note that this is based partly on the Doomsday Argument discussed in section 3.

[11] Note that Bostrom does not provide a time limit.

> (vii) And the Doomsday Clock, maintained by the *Bulletin of the Atomic Scientists*, is currently set to 2 minutes before midnight (or doom). Only in 1953, after the US and Soviet Union detonated thermonuclear bombs, was the minute hand this close to striking twelve (Mecklin 2018).[12]

If one takes these estimates seriously—and I will present reasons below for doing so—they imply that the average American is literally *thousands of times* more likely to encounter a existential disaster than, say, to die in an air and space transport accident. More specifically, the FHI survey suggests that the average American is at least 1,500 times more likely to perish in a human extinction event than a plane crash, and Rees's estimate implies that the average American is nearly 4,000 times more likely to encounter the collapse of civilization than die in an aviation mishap. If this is even remotely accurate, a child born today has a nontrivial chance of living to witness an existential catastrophe of some sort (Torres 2017a, 2017b). Wells (2009) makes a similar point when he asks, "which is more likely: that your house burns down, or you perish in a global cataclysm? If you live in an ordinary urban house with a fire station at a normal distance, and if you have no implacable enemy, then death in a global disaster is more likely."

These estimates are also consistent with warnings made by a number of leading intellectuals, philosophers, and scientists around the world. For example:

> (viii) The late Stephen Hawking writes in a *Guardian* op-ed "that we are at the most dangerous moment in the development of humanity" (Hawking 2016), subsequently suggesting that our species has about 100 years to leave planet Earth "or die" (Fecht 2017).
> (ix) Richard Posner judges the near-term chance of extinction to be "significant," adding that "human extinction is becoming a feasible scientific project" (Posner 2004).
> (x) Michio Kaku argues that "the danger period is *now*. Because we still have the savagery. We still have all the passions. We have all the sectarian fundamentalist ideas circulating around. But we also have nuclear weapons. We have chemical, biological weapons capable of wiping out life on Earth" (Kaku 2011).[13]
> (xi) The great polymath Noam Chomsky estimates that the risk of human annihilation is currently "unprecedented in the history of *Homo sapiens*" (Lombroso 2016).
> (xii) Max Tegmark contends that "it's probably going to be within our lifetimes … that we're either going to self-destruct or get our act together" (Harris 2018).
> (xiii) Paul Ehrlich prognosticates that the collapse of civilization as a "near certainty" in the coming decades if humanity continues destroying the natural world at current rates (Carrington 2018; Ehrlich and Ehrlich 2009).
> (xiv) And Ingmar Persson and Julian Savulescu (2012) argue that "our present situation is so [existentially] desperate" that society should consider taking extreme measures, such as implementing invasive surveillance systems and providing moral bioenhancement options to its citizens, to obviate what they call "Ultimate Harm," which would "render worthwhile life forever impossible."[14]

---

[12] Others have made similar conjectures. For example, Frank Fenner speculated in 2010 that "humans will probably be extinct within 100 years" (Edwards 2010); Neil Dawe says that "he wouldn't be surprised if the generation after him witnessed the extinction of humanity" (Jamail 2013); James Lovelock has conjectured that the global population will be reduced by "about 80%" as of 2100; Mayer Hillman declared in 2018 that "we're doomed. The outcome is death, and it's the end of most life on the planet because we're so dependent on the burning of fossil fuels" (Barkham 2018); and Benjamin Todd, writing for 80000 Hours, states that "overall, we think the risk is likely over 3%" (Todd 2017). In contrast, Richard Gott uses statistical analyses to estimate that humanity will go extinct between 5,100 and 7.8 million years from now, and Bruce Tonn (2009) conjectures that "the probability of human extinction is probably fairly low, maybe one chance in tens of millions to tens of billions, given humans' abilities to adapt and survive."

[13] Italics added.

[14] It is worth adding that Roy Scranton, who authored *Learning to Die in the Anthropocene*, argues that humanity is ultimately "doomed" because of climate change and that, therefore, living in the Anthropocene is about "palliative care" for civilization, which is is in "hospice" (see Scranton 2013; FOP 2015). See also Elon Musk's estimate of the immense risks associated with superintelligence in subsection 2.3.

So, to summarize, the number of risk scenarios and the probability of any given scenario occurring are both increasing, and many leading intellectuals maintain that humanity is in a uniquely vulnerable situation. This is the other side of the futurological coin that Pinker and his New Optimist colleagues habitually fail to seriously consider—empirical trends that imply that residents of spaceship Earth are are (much) greater risk today of annihilation this century than ever before, *even as* levels of global violence have declined and our circles of moral concern have expanded.

The following three subsections examine each Great Challenge, namely, environmental degradation, the democratization of science and technology, and machine superintelligence, in turn. The final section explores a number of complicating issues, such as the "ignorance explosion" and "toxic masculinity."

Section 2: The Great Challenges

The threat of the apocalypse will be with us for a long time.—Robert Oppenheimer

2.1 Environmental Degradation

Let's begin with the proximate chemical cause of climate change, namely, carbon dioxide (CO2). The highest concentration of CO2 in the ambient air over the past 400 million years was ~300 parts per million (ppm) and our ancestors evolved with concentrations between 180 and 280 ppm. Yet studies show that there could be upwards of 1,000 ppm of CO2 by the end of this century (see Torres 2017a). The resulting climatic changes, according to the Intergovernmental Panel on Climate Change (IPCC), will be "severe," "pervasive," and "irreversible" (IPCC 2014). Such changes include extreme weather events, megadroughts lasting decades, devastating coastal flooding, sea-level rise, melting glaciers and the polar icecaps, desertification, deforestation, food supply disruptions, infectious disease outbreaks, mass migrations, and heat waves that surpass the 95 degree wet-bulb threshold for human survivability, meaning that even if one were naked in the shade in front of a giant fan, death would still be inevitable (Willett and Sherwood 2012; Torres 2016, 2017a). In fact, a large 2017 study notes that about 30 percent of the global population is exposed to "lethal heat events" for 20 or more days a year. But if greenhouse emissions continue to grow, approximately 74 percent will be exposed to this "deadly threshold," and even if humanity drastically reduces its emissions, the percentage will still rise to about 50 (Mora et al. 2017). As another study reports, between 20 and 30 percent of the planet will undergo aridification if the global mean temperature rises to 2 degrees Celsius (Park et al. 2018). Even more bizarrely, scientists project that lightning strikes will increase 50 percent by 2100, allergy seasons will last longer and become more intense, and Earth's tilt and rotational speed will change (see Torres 2017a). The "threat-multiplying" or "conflict-multiplying" consequences of climate change will also severely destabilize governments and produce social upheaval, economic turbulence, conflicts, and terrorism (Torres 2017a).

In fact, the hottest 18 years on record have all occurred since 2000, with one exception, viz., 1998. The year 2016 holds the record, with 2017 in second place according to the National Oceanic and Atmospheric Administration (NOAA), followed by 2015. But unlike 2015, 2017 was not an El Niño year, thus making it "the hottest year without an El Niño by a wide margin" (Nuccitelli 2018). And new research has linked the heat of 2016 with extreme heat waves in Asia that killed over 500 people, and an anomalous warm "blob" off the West Coast of the US that could have contributed to strange weather as far away as the East Coast. Such data spurred the NASA astronaut Mark Kelly to describe 2017—during which little action was taken to curb climate change after the Republican Party, arguably the only major political party in the world that still questions climate science, rose to prominence—as "an unequivocal disaster for the future of the planet" (Kelly 2017).

But "curbing climate change" might require more than simply reducing our collective carbon footprint. As Jacob Haqq-Misra and his colleagues write in an analysis of the Malthusian problem of populations growing faster than their food supply finds

> that, even if greenhouse gas emissions are mitigated, growth in human civilization's energy use will thermodynamically continue to raise Earth's equilibrium temperature. If current energy consumption trends continue, then ecologically catastrophic warming beyond the heat stress toler-

> ance of animals … may occur by ~2200-2400, independent of the predicted slowdown in population growth by 2100 (Haqq-Misra at al. 2017).[15]

Anthropogenic CO2 is also causing ocean acidification, which has resulted in significant marine biodiversity loss. In fact, the *rate* of ocean acidification today is probably faster than the rate at which it occurred 250 million years ago during the "Great Dying," or Permian-Triassic extinction, that eliminated 95 percent of all species on Earth. Whereas 2.4 gigatons of carbon were injected into the atmosphere per year during this extinction (much of which ended up in the oceans), scientists estimate that civilization injects about 7.6 gigatons of carbon *more* into the atmosphere per year (Hand 2016). Scientists have also identified nearly 550 aquatic dead zones, or hypoxic bodies of water, around the world, with the largest being slightly smaller than the sizes of New Hampshire, Vermont, and Maryland added together (see Torres 2016, 2017a).[16]

One consequence of ocean acidification and warmer waters is coral bleaching. Right now, about half of the world's coral reefs have become underwater ghost towns and about 90 percent of them are projected to die by 2050 (Becatoros 2017). In fact, "coral reefs provide homes and nursery ground to many fish species," with "about one-third of all saltwater fish species [living] at least part of their lives on coral reefs" (DoW 2018). Thus, it is perhaps unsurprising that, according to another study, if current trends are extrapolated into the future, there will be (virtually) no more wild-caught seafood by 2048 (Worm et al. 2006). Even more, some researchers have speculated that ocean warming could interfere with the photosynthesis of phytoplankton, which currently provides "about two-thirds of the planet's total atmospheric oxygen" (Sekerci and Petrovskii 2015; SD 2015). If this were to occur, it could lead to a catastrophic decline in atmospheric oxygen levels, thus resulting "in the mass mortality of animals and humans," as the authors put it.

But the diminution of biological diversity is a problem far larger than this. According to the Global Biodiversity Outlook (GBO-3) report from 2010, the total population of wild vertebrates between the Tropic of Cancer and the Tropic of Capricorn fell by a staggering 59 percent in only 36 years, from 1970 to 2006. (The taxon of vertebrates includes mammals, birds, fish, reptiles, and amphibians.) The report also found that vertebrates in freshwater environments declined by 41 percent, farmland birds in Europe declined by 50 percent since 1980, birds in North America declined by 40 percent between 1968 and 2003, and about 25 percent of all plant species—the foundation of the food chain—are currently "threatened with extinction" (see Torres 2016, 2017a).[17] Similarly, the 2016 Living Planet Report states that the global abundance of wild vertebrates declined by an incredible 58 percent between 1970 and 2012, and we could witness a decline of 2/3rds by 2020 (WWF 2014), whereas the 2018 Living Planet Report concludes that, "on average, we've seen an astonishing 60% decline in the size of populations of mammals, birds, fish, reptiles, and amphibians in just over 40 years." Other studies have found that 19 percent of all reptile species, 50 percent of freshwater turtles (Böhm et al. 2013), and ~60 percent of the world's primates are under threat, while the populations of ~75 percent are declining (Estrada et al. 2017). And "the most important insect that transfers pollen between flowers and between plants," namely, the honey bee, is struggling as a result of *colony collapse disorder* (Torres 2017a). This has implications for agricultural production, an especially unsettling fact given that one study estimates that we will need to produce more food in the coming 50 years than we have produced in our entire history so far (Potter 2009). In fact, the UN has calculated the future human population size based on "low" and "high" variants. The former gives a population of 7.3 billion whereas the latter gives one of a staggering 16.5 billion (UN 2017). Complicating matters even more, soil erosion is reducing the annual crop yield by 0.3 percent, meaning that "at this rate, we will have lost 10 percent of soil productivity by 2050"—about the same loss that global warming is expected to cause (Kuhlemann 2018).

---

[15] On a somewhat related note, a recent article in *Nature Climate Change* calculates that "Bitcoin usage, should it follow the rate of adoption of other broadly adopted technologies, could alone produce enough CO2 emissions to push warming above 2 °C within less than three decades" (Mora et al. 2018).

[16] Thanks to Robert Diaz for this number (personal communication).

[17] Note that people often don't consider how human activity might be affecting plant life, a phenomenon that scientists have dubbed "plant blindness."

Statistics such as these have led numerous scientists to worry about the possibility of rapid changes to the biosphere that could imperil civilization. For example, a study from 2012 argues that civilization could be barreling toward a planetary-scale "state shift," which could precipitate "substantial losses of ecosystem services required to sustain the human population." If a sudden, irreversible, and catastrophic collapse of the global ecosystem were to occur, it would likely produce "widespread social unrest, economic instability, and loss of human life" (Barnosky et al. 2012). This comports with a highly influential report authored by nearly 30 scientists, including several Nobel laureates, which identifies nine Earth-system processes associated with *planetary boundaries*, including (i) climate change, (ii) ocean acidification, (iii) stratospheric ozone depletion, (iv) atmospheric aerosol loading, (v) biogeochemical flows (i.e., phosphorus and nitrogen cycles), (vi) global freshwater use, (vii) land-system change, (viii) rate of biodiversity loss, and (ix) chemical pollution. Together, these demarcate a "safe operating space for humanity" in which sustainable development must proceed or else risk disaster. As the authors write,

> anthropogenic pressures on the Earth System have reached a scale where abrupt global environmental change can no longer be excluded. … Transgressing one or more planetary boundaries may be deleterious or even catastrophic due to the risk of crossing thresholds that will trigger non-linear, abrupt environmental change within continental- to planetary-scale systems.

The report adds that "humanity has already transgressed three planetary boundaries: for climate change, rate of biodiversity loss, and changes to the global nitrogen cycle," meaning that we are now vulnerable to global environmental transitions that could unfold rapidly and severely harm civilization (Rockström et al. 2009).

There are two ways of measuring biodiversity: the size of populations and the number of species. So far, we have focused primarily on the former; but data about the latter paints and even more dire picture. Today, the biological extinction rate is between 100 and 1,000 times higher than the normal "background" extinction rate, and "99 percent of currently threatened species are at risk from human activities" (Center 2018). The result is that we have entered the sixth mass extinction event in the 3.8 billion year history of Earth-originating life: the "Anthropocene extinction," which could be our greatest legacy on the planet. The reality of this slow-motion catastrophe is no longer controversial. As a 2015 study in *Science Advances* reports, even the most optimistic assumptions about the background rate of species losses and the current rate of vertebrate extinctions imply an extinction event (see Torres 2017a); the authors write that the evidence clearly confirms "an exceptionally rapid loss of biodiversity over the last few centuries, indicating that a sixth mass extinction is already under way" (Ceballos 2015). Thus, we may begin talking about the "Big Six" instead of the "Big Five," which denotes the previous five mass extinction events in life's biography—the most recent one being the extinction of most of the dinosaurs some 66 million years ago.

Incidentally, the term "Anthropocene" is of quite recent provenance, having been proposed by scientists to denote a new geological epoch whose physical signatures include major climatic and biospheric disruptions and the distribution of artificial radionuclides around the world from thermonuclear testing. As Jennifer Jacquet (2017) notes in an article about this epoch, "not since cyanobacteria has a single taxonomic group been so in charge. Humans have proven we are capable of seismic influence, of depleting the ozone layer, of changing the biology of every continent." The question now is "whether we are prepared for this kind of control."

These phenomena pose myriad direct threats to the perpetuation and flourishing of our species. Many civilizations throughout history, including the Mayan and Rapa Nui civilizations, have of course collapsed due to environmental degradation caused by deforestation, pollution, overfishing, and other forms of ecological destruction (see Diamond 2005). In fact, a NASA-funded study from 2014 uses a mathematical model—the "human and nature dynamical model," or "HANDY"—to show that the overexploitation of natural resources, along with wealth inequality, can precipitate the collapse of advanced civilizations (Motesharrei et al., 2014). The authors warn that contemporary civilization is dangerously close to bringing about its own collapse as a result of both phenomena—that is, we are dangerously close to falling victim to what Daniel O'Leary (2006) calls a "progress trap." Indeed, not only is the environment degrading, with the worst effects impacting impoverished countries the most, an increasingly urgent issue of "climate justice," but a 2018 World Inequality Report found that "the top 0.1 percent has captured

as much growth [in wealth] as the bottom half of the world adult population since 1980" (Alvaredo 2018). And whereas the most affluent person in the world, Jeff Bezos, has a net worth of $105 billion, almost half of the world's population—3.8 billion human beings—survives on less than 2.50 USD per day.[18]

Yet another clarion warning that humanity could be barreling toward disaster comes from a short paper titled "World Scientists' Warning to Humanity: A Second Notice." Signed by *over 15,000 scientists*, it observes that "humanity has failed to make sufficient progress in generally solving … foreseen environmental challenges, and alarmingly, most of them are getting far worse." The conclusion is that

> to prevent widespread misery and catastrophic biodiversity loss, humanity must practice a more environmentally sustainable alternative to business as usual. This prescription was well articulated by the world's leading scientists 25 years ago [when the "first notice" was published], but in most respects, we have not heeded their warning. Soon it will be too late to shift course away from our failing trajectory, and time is running out. We must recognize, in our day-to-day lives and in our governing institutions, that Earth with all its life is our only home (Ripple et al. 2017).

This is far from an exhaustive survey of the environmental challenges facing humanity today. Indeed, I have not discussed phenomena like the "greening" of the Antarctic, the protracted photodegradation period of plastics, the "zombie pathogens" that are emerging from thawed permafrost, the forward creep of Overshoot Day, and the possibility of rapid increases in atmospheric methane that could, potentially, initiate a runaway greenhouse effect (the "clathrate gun hypothesis"). Suffice it to say that environmental degradation is a significant, urgent, and ineluctable problem that humanity will need to solve if we wish to survive on this pale blue dot—Planet A. Or, to quote the former Romanian environment minister: "I hope we aren't the first species to document our own extinction" as a result of environmental degradation.

2.1 The Democratization of Science and Technology

Prior to 1945, no single actor had the capacity to unilaterally destroy the world; after 1945, two state actors acquired this capacity. Today, techno-developmental trends suggest that the number of state and nonstate actors with this capacity is increasing and will continue to increase this century. There are three aspects, in particular, of emerging technologies—most notably, biotechnology, synthetic biology, nanotechnology, and artificial intelligence[19]—that one must understand to appreciate the danger. These are:

> (i) *Use-Flexibility*. Emerging technologies are *dually usable*, meaning that the very same artifact can be used for both beneficial and harmful ends. For example, centrifuges that can enrich uranium for nuclear power plants can also enrich it for nuclear weapons; a laboratory that could find a cure for Ebola could also be used to weaponize this already-deadly virus; and so on.
> (ii) *Capability*. Emerging technologies are increasingly *powerful*, thus enabling actors to manipulate and rearrange the physical world in unprecedented ways. This trend appears to be unfolding at an exponential or superexponential rate, along the lines of Moore's law, Huang's law,[20] the Carlson curve, Dennard scaling, Keck's law, Kryder's law, and other trends subsumable under the Kurzweilian "Law of Accelerating Returns" (Kurzweil 2005).
> (iii) *Democratization*. Emerging technologies are increasingly *accessible* to small groups, lone wolves, and "lone wolf packs" (see Pantucci 2011). Examples include CRISPR/Cas-9, digital-to-biological converters, base editing, USB-powered DNA sequencers, SILEX (i.e., the separation of uranium isotopes by laser excitation), as well as anticipated future artifacts like nanofactories, which could enable state and nonstate actors to manufacture huge arsenals of advanced weapon-

---

[18] Although others have suggested that Vladimir Putin is actually the wealthiest.

[19] There are many other technologies that could be added to this list, such as the those involved in stratospheric geoengineering and deflecting asteroids.

[20] This refers to the development of GPUs. During a 2018 talk, Jensen Huang claimed that "there's a new law going on, a supercharged law" (Perry 2018). For example, some GPUs are 25 times faster than only 5 years ago, which is 15 times faster than if Moore's law had been governing their development.

> ry; autonomous nanobots that could target specific people, races, or species; lethal autonomous drones—e.g., "slaughterbots" (Russell et al. 2018[21])—that are programmed to wipe out entire cities; and even asteroid deflection spacecraft that future nanotechnology could make available to terrorist organizations or individuals. We should also expect that metamaterial invisibility cloaks, self-guided bullets, cognitive enhancements like nootropics and brain-machine interfaces, exoskeletons, robot soldiers, direct-energy weapons (DEWs) like laser and particle beam weapons, and mind-control/mind-reading technologies will further complicate the situation.

It is important to recognize that the technologies listed above could also bring about truly marvelous improvements in the human condition by eliminating disease, reversing aging, and perhaps enhancing human morality (Kurzweil 2005; Diamandis 2014; Persson and Savulescu 2012).[22] Yet the property of (i) entails that these very same inventions will simultaneously engender profound risks to our collective well-being and survival. As Robert Oppenheimer observed as far back as 1956, "it is not new that knowledge brings power, and that among the powers may be the power to do evil. In modern science there is much such knowledge."

Furthermore, given a conception of technology use—or at least the use of certain types of technologies—as instantiating an ontology of *agent-artifact couplings*, we can, focusing on the agent side of the dyad, distinguish between two subtypes of agential risk: error and terror (see Rees 2003; Torres 2017a). Taking these in reverse order, I have recently outlined a quadripartite typology of human agents who are prime candidates for intentionally bringing about an existential catastrophe once the means become sufficiently available. These are: (1) *Apocalyptic terrorists*, e.g., religious extremists who believe that the world must be destroyed to be saved; (2) *misguided moral actors*, e.g., "radical negative utilitarians" (RNUs) who advocate annihilation to eliminate suffering; (3) *radical ecoterrorists*, e.g., deep ecology extremists who believe that the biosphere would be better off without *Homo sapiens*; and (4) *idiosyncratic actors*, e.g., rampage shooters who have wished to kill as many people as possible before dying (Torres 2017b, 2017c). Consider a few quotes from actual individuals in categories (3) and (4) to get a sense of how dangerous they could be in the emerging threat environment outlined above: First, in an issue of the *Earth First! Journal*, an anonymous author writes that

> contributions are urgently solicited for scientific research on a species specific virus that will eliminate Homo shiticus from the planet. Only an absolutely species specific virus should be set loose. Otherwise it will be just another technological fix. (Dye 1993).

This pro-omnicide view is unsettlingly common among individuals within the most extreme fringe of the biocentric environmentalist movement. For example, the Toronto-based Gaia Liberation Front (GLF) states that its "mission is the total liberation of the Earth, which can be accomplished only through the extinction of the Humans as a species," to which it adds that this could be accomplished involuntarily through some global-scale catastrophe like an engineered pandemic (Torres 2017a, 2017b, 2017c). As for rampage shooters driven by idiosyncratic motives, the mastermind behind the 1999 Columbine school massacre, Eric Harris, wrote in his journal, "if you recall your history the Nazis came up with a 'final solution' to the Jewish problem. Kill them all. Well, in case you haven't figured it out yet, I say 'KILL MANKIND' no one should survive." To this he added, "I think I would want us to go extinct," "I have a goal to destroy as much as possible … I want to burn the world," and "I just wish I could actually DO this instead of just DREAM about it all" (Torres 2017c). And finally, the incel rampage shooter Elliot Rodger declared in a video recorded just days before his attack,

> I hate all of you. Humanity is a disgusting, wretched, depraved species. If I had it in my power, I would stop at nothing to reduce every single one of you to mountains of skulls and rivers of blood. And rightfully so. You deserve to be annihilated. And I'll give that to you (quoted in Garvey 2014).

---

[21] See also SAW 2017.

[22] Although see Torres 2017 for criticism of moral bioenhancement from the "agential risk" perspective.

Many individuals like Harris and Rodger have suffered from sociopathy (or psychopathy), which affects between 1 and 4 out of every 100 people. This means that there are ~300 million sociopaths in the world today and there will be ~372 million by 2050, if the global population rises to 9.3 billion (Stout 2005; Torres 2017a, 2017c). While not all sociopaths are sadistic or violent, they do comprise a disproportionate percentage of the prison population—about 20 percent (Torres 2017a, 2017c). Thus, there is a growing pool of individuals with personality disorders from which future "idiosyncratic actors" could emerge. The point of these brief examples is to underline that there *really are* people who have (a) demonstrated a willingness to engage in catastrophic violence by perpetrating mass shootings, and (b) entertained omnicidal ideations, expressed either in private journals or public conversations. As the trends of (ii) and (iii) continue, such agents will pose an increasingly significant existential danger, a proposition further bolstered by analyses that lone wolf attacks have steadily increased every decade since 1970 (Spaaij 2010), and single terrorist attacks have become incrementally more lethal (Miller 2015; Jenkins et al. 2016).

With this in mind, let's quantify the hypothetical threat level that terror agents could pose, I propose the relatively crude, but nonetheless useful, *AW formula*, where "AW" stands for the necessary and sufficient conditions of "able and willing." By definition, all terror agents would be willing to destroy the world if only they were able, i.e., the probability of such individuals pressing a "doomsday button," if within reach, would be ~100 percent. Fortunately, the "A" variable right now has a low probability of being satisfied—that is to say, it is currently difficult for malicious individuals and terrorist organizations to acquire technologies capable of global-scale destruction.[23] But how low does this probability need to be to ensure the safety of civilization? For the sake of illustration, let's posit that there are 1,000 terror agents in a population of 10 billion and that the probability per decade of *any one* of these individuals gaining access to world-destroying weapons (thus satisfying the "A" factor) is only 1 percent. What overall level of existential risk would this expose the entire population to? It turns out that, given these assumptions, the probability of a doomsday attack per decade would be a staggering 99.995 percent. One gets the same result if the number of terror agents is 10,000 and the probability of access is 0.1 percent, or if the number is 10 million and the probability is 0.000001. Now consider that the probability of access may become *far greater* than 0.000001—or even 1—percent, given the trend of (iii), and that the number of terror agents could exceed 10 million, which is a mere 0.1 percent of 10 billion. It appears that an existential strike could be more or less inescapable.

But the situation may be far worse than this because of agential error. This claim follows in part from the fact that the class of terror agents is a subset of the class of error agents: all of the former instantiate the latter (given universal human fallibility), whereas relatively few of the latter instantiate the former (given that most people don't want to destroy the world). Here we can use the *EA formula* to estimate the danger posed by error agents in the future. The acronym "EA" stands for "error-proneness and able." Whereas above we asked about the probability of any given terror agent within a population gaining access to world-destroying weapons, here we can ask about the probability of error given some population of non-terror agents with *ex hypothesi* access to powerful dual-use technologies. Although one might assume that this probability would be extremely low, the history of mistakes, even in highly regulated government or university laboratories, that have had real-world consequences is quite sobering. For example, the 1977 swine flu outbreak was probably the result of a laboratory leak (Zimmer and Burke 2009); a government report specifies more than 1,100 laboratory blunders involving hazardous biomaterials between 2008 and 2012; in 2014, some 75 scientists at the Centers for Disease Control (CDC) "may have been exposed to live anthrax bacteria after potentially infectious samples were sent to laboratories unequipped to handle them"; and "a CDC lab accidentally contaminated a relatively benign flu sample with a dangerous H5N1 bird flu strain that has killed 386 people since 2003" (see Torres 2016, 2017a). Such accidents could become even more common as the global community of professional scientists grows—according to one count, roughly 1.5 million scientists published articles that are indexed under "genetic techniques" between 2008 and 2015 (Sotos 2017). Adding to this statistical risk, the biohacker movement is also gaining momentum, as mail-order CRISPR kits and other DIY bio-lab equipment become more

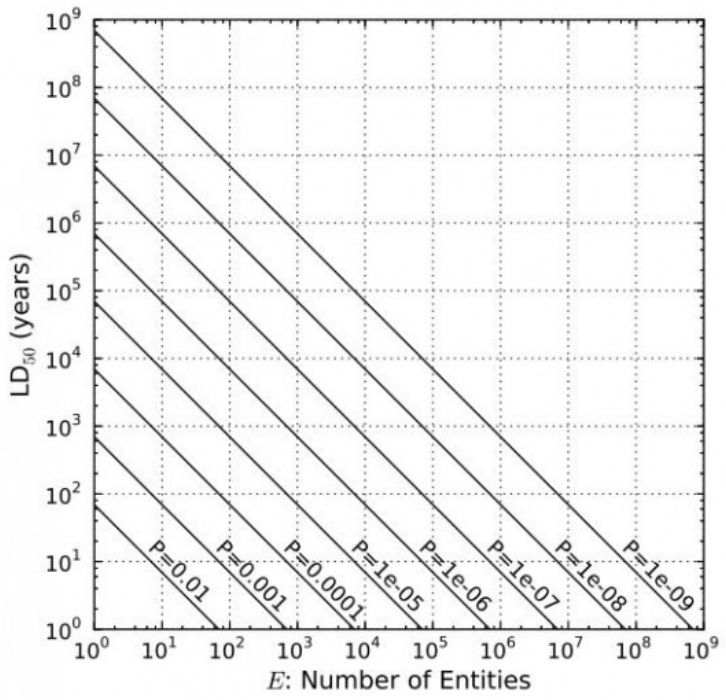
[23] Indeed, some of these technologies have not yet been developed.

affordable to amateur hobbyists. Thus, the overall probability that *someone somewhere sometime* will make a catastrophic mistake may not be small, and is probably growing.[24]

This has implications exactly similar to those above. To drive home the point, let's consider another calculation that reveals just how small the probability needs to be for a catastrophe to be more or less certain. First, imagine a world in which only 500 of 10 billion non-terror agents—that is, a mere 0.000005 percent of the population—have access to world-destroying technologies. If each agent has a mere 0.01 chance of *accidentally* initiating a doomsday disaster per decade, the probability of self-annihilation would be an astonishing 99.3 percent (Torres 2017a, 2017b). Doom becomes even more certain if all 10 billion individuals (a) have access to world-destroying technologies, and (b) have a negligible 0.000000001 chance of accidentally pressing a doomsday button. Once again it appears that the trend of (iii), plus the trend of (ii) and the property of (i), may pose an exceptionally high risk to our collective survival. As Rees (2003) puts this general point,

> If there were millions of independent fingers on the button of a Doomsday machine, then one person's act of irrationality, or … one person's error, could do us all in. … Disastrous accidents (for instance, the unintended creation or release of a noxious fast-spreading pathogen, or a devastating software error) are possible even in well-regulated institutions. As the threats become graver, and the possible perpetrators more numerous, disruption may become so pervasive that society corrodes and regresses. There is a longer-term risk even to humanity itself.[25]

Incidentally, John Sotos (2017) has independently proposed a similar model that yields comparably pessimistic results. Focusing on biotechnology, he calculates that a 1 in 100 chance of only a few hundred



[24] Or, as Bostrom (2014) gently puts it, "some little idiot is bound to press the ignite button."

[25] Another quote that I often use comes from James Fearon: "A friend of mine, a journalist, quips that we seem to be heading in the direction of a world in which every individual has the capacity to blow up the entire planet by pushing a button on his or her cell phone. … How long do you think the world would last if five billion individuals each had the capacity to blow the whole thing up? No one could plausibly defend an answer of anything more than a second. Expected life span would hardly be longer if only one million people had these cell-phones, and even if there were 10,000 you'd have to think that an eventual global holocaust would be pretty likely. Ten thousand is only two millionths of five billion" (quoted in Torres 2017a).

agents releasing a species-destroying pathogen—whether for error or terror reasons—yields virtually inevitable doom within ~100 years. Even more, if the total number of agents capable of inflicting global-scale harm rises to 100,000, the probability of any one person releasing such a pathogen must be less than 1 in 10^9 for civilization to survive a millennium (see MIT 2017). Sotos derives these figures from a model in which "the projected lifetime of a civilization (LD50) depends inversely on the number of people, or entities, capable of destroying it (E) and the probability per year that one of them will (P)" (Figure 1). The term "LD50" refers to the "lethal duration 50" of a civilization, indicating "the number of years, under a given E and P, before civilization's accumulated probability of being uncommunicative … is 50%." In fact, Sotos concludes that his calculations supply "the quantitative 24 orders-of-magnitude winnowing required of a Great Filter," given that the visible universe contains "≈ 10^{24} stars and their planets" and "only Earth shows evidence of intelligent life." Thus, if "civilizations universally develop advanced biology, before they become vigorous interstellar colonizers, the model provides a resolution to the Fermi paradox" (Sotos 2017). With respect to the genealogy of ideas, Sotos draws from and elaborates the work of Joshua Cooper, who argues that any species capable of colonizing space will have (a) a very large population, and (b) sophisticated knowledge about its own biochemistry. Cooper provides compelling reasons for this claim, which I won't here recapitulate. Suffice it to say that these considerations, he argues, "provide a neat, if profoundly unsettling, solution to Fermi's paradox" (Cooper 2013).

There is a sense in which it appears that the hackneyed phrase "it's a matter of when rather than if" applies to this general phenomenon. The demographic of potential perpetrators is growing too large and quickly for humanity—perhaps the dumbest species capable of building world-destroying weapons—to escape a disaster.[26]

2.3 Machine Superintelligence

Many existential risk scholars, including the present author, believe that superintelligence probably constitutes the greatest known threat to our long-term survival (Bostrom 2014; Yampolskiy 2016; Torres 2017a). The reason, in brief, is this: (a) while it would be confused to say that "intelligence is power," it would not be inaccurate to assert that "intelligence yields power" (Torres 2017d). Humanity provides a compelling example: We are the most dominant creature on the planet not because of our ability to smash, strike, throw, bite, or scratch, but because of our superior encephalization quotient (EQ). Thus, a superintelligence would be superpowerful, its efferent interfaces being any technical artifact or system within electronic reach. And (b) there is no known solution to the *control problem*, or the conundrum of figuring out how to sufficiently align the value system of a superintelligent machine with our own value system, whatever that is. Here we can quickly dismiss criticisms of "Friendly AI" safety engineering like those articulated by Pinker:

> AI dystopias project a parochial alpha-male psychology onto the concept of intelligence. They assume that superhumanly intelligent robots would develop goals like deposing their masters or taking over the world. But intelligence is the ability to deploy novel means to attain a goal; the goals are extraneous to the intelligence itself. Being smart is not the same as wanting something. … It's telling that many of our techno-prophets don't entertain the possibility that artificial intelligence will naturally develop along female lines: fully capable of solving problems, but with no desire to annihilate innocents or dominate the civilization (Pinker 2015).

This falls victim to what Max Tegmark (2018) calls one of "the top myths of advanced AI." The worry isn't that superintelligence could transmogrify into an evil, malicious, domineering alpha-male terminator, but rather that its goals could be misaligned with ours *and* it could be competent enough to achieve those

---

[26] Here we might also note Persson and Savulescu's (2012) insight that "it is comparatively easier to harm than to benefit." This is why, they argue, "misuses of scientific discoveries pose such an alarming threat. For the easiness to harm is magnified as our powers of action increase through technology, the power to harm always keeping its clear lead over any expanding power to benefit."

goals.[27] More precisely, we can dissect the danger posed by superintelligence into the following six theses:

> (i) *Orthogonality thesis*. This states that the space of combinatorial possibility with respect to intelligence and final goals (or "values") is in principle unconstrained by either variable. In other words, virtually any level of intelligence can be combined with virtually any final goal, meaning there is no contradiction with a genuinely superintelligent machine that only values playing tic-tac-toe, writing and rewriting Shakespeare's plays, proving Goldbach's conjecture, maximizing the total number of paperclips in the universe, or counting blades of grass on Harvard's campus.[28]
> (ii) *Instrumental convergence thesis*. This states that agents with a wide range of final goals are all likely to converge upon a finite number of intermediate subgoals. For example, acquiring unlimited resources—including the atomic particles that comprise our bodies—would be instrumentally desirable for achieving all of the goals listed above. The same goes for ensuring its continued existence, preventing changes to its utility function, augmenting its intelligence, and acquiring a complete theory of the physical universe.
> (iii) *Complexity of value thesis*. Our values, whatever they are exactly, have a high Kolmogorov complexity. That is to say, they cannot be reduced to a simple codifiable list of prescriptions and proscriptions.
> (iv) *Fragility of value thesis*. It appears that successfully loading *most* of our values into an artificial intelligence may be insufficient to guarantee proper value alignment; we will need to upload *all* of them. As Sandberg puts it, paraphrasing others, "getting a goal system 90 percent right does not give you 90 percent of the value, any more than correctly dialing 9 out of 10 digits of my phone number will connect you to somebody who's 90 percent similar to me" (Sandberg 2017).
> (v) *Relative speed thesis*. The electrical potentials within computer hardware can transfer information about a million times faster than the action potentials within our brains. This means that a single minute of objective time would equal about 2 years of subjective time for a computer-emulated human brain. Thus, whereas it takes the average PhD student 8.2 years or so to obtain a degree, an uploaded mind could achieve this in a matter of 4.3 minutes. The "speed of thought" differential between humans and computers could give the latter an immense strategic advantage over humanity. And finally,
> (vi) *Rapid capability gain thesis*. This is associated with the instrumental value of "cognitive enhancement": For nearly any given final goal, being smarter would facilitate achieving that goal. But a machine recursively improving itself could initiate a positive feedback loop that produces an *intelligence explosion*, thus resulting in a superintelligence that is, perhaps, more intelligent than humans to the same extent that humans are more intelligent than dung beetles.

There are several additional issues that render the threat posed by superintelligence even more formidable. For example, theses (v) and (vi)—along with the "intelligence yields power" truism—suggest that humanity may have *one and only one chance* to get the problems of (iii) and (iv) *exactly right*. There probably won't be the option of scrapping a failed superintelligence project because it is about to destroy us and starting over. Even more, humanity may not have much time to solve these high-stakes problems: a recent survey of AI experts places the median probability estimate for the creation of human-level AI before 2075 at 90 percent (Müller and Bostrom 2014). If the method used to create such an AI is what David Chalmers (2010) describes as "extendible," then there are reasons for believing that a superhuman-level AI will follow shortly after human-level AI—meaning that there is an unsettling chance that a superintelligence will join humanity by 2100.[29] Another survey reports that 75 percent of respondents—all fellows of the Association for the Advancement of Artificial Intelligence—believed that superintelligence would someday become a reality, with 7.5 percent expecting this to happen in the next 10 to 25 years and the remainder believing that it will happen in more than 25 years (Etzioni 2016). Yet another calculates the

---

[27] For a more detailed critique of Pinker, see Torres 2018.
[28] I borrow a version of the grass-counting example from Rawls 1999.
[29] That is, insofar as expert opinion is credible; see Armstrong and Sotala 2012.

"the aggregate forecast [of experts] gave a 50% chance of HLMI occurring within 45 years and a 10% chance of it occurring within 9 years," where "HLMI" is acronymous for "high-level machine intelligence," which "is achieved when unaided machines can accomplish every task better and more cheaply than human workers" (Grace et al. 2017). This survey also found that 29 percent of experts believe that an intelligence explosion is either "likely" or "highly likely."[30]

So, it appears that by the end of this century—if not within a few decades, if not within a few years—we may need to have solved the *philosophical problem* of what exactly our values are as well as the *technical problem* of how to load them into a machine (Yudkowsky 2008b). The first could ultimately be insoluble. Indeed, not even professional ethicists can agree about the most basic practical, normative, and metaethical issues (see Bourget and Chalmers 2014). Second, while programming simple goals like "manufacture 1,000 paperclips" is relatively easy, it is far more difficult to encode abstract human concepts like "happiness" and "well-being" in "the AI's programming language, and ultimately in primitives such as mathematical operators and addresses pointing to the contents of individual memory registers" (Bostrom 2014). There is also the issue of *value drift*, or the possibility that we get everything 100 percent right with respect to the control problem on the first go, yet we fail to ensure that the values loaded into the AI are sufficiently stable, thus resulting in axiological mutations that accumulate over time to gradually turn a "friendly" algorithm "unfriendly," given (ii) above. The flip problem is *value ossification*, or the possibility that we overcome value drift but later realize that the values we loaded into the AI are suboptimal in one or more crucial ways, yet we are unable to modify them. Consider how much "our values" have changed over time: for example, cat burning was morally acceptable in eighteenth-century France. Extrapolating this into the future, we could develop values at time T2 that supplant our previous values at T1, making it problematic that our T1 values have been "locked in" to the superintelligence.

It is problems like these that lead Nick Bostrom to suggest that we should recognize the "default outcome" of creating a superintelligence to be "doom" (Bostrom 2014). Elon Musk echoes this sentiment, declaring that superintelligence constitutes a "fundamental risk to the existence of human civilization" and that we have "a 5 to 10 percent chance" of *avoiding* an existentially bad outcome (Gohd 2017). Given that progress in computer science will continue to accelerate in the coming years and decades—that is, in the absence of a defeater like an existential catastrophe—this risk appears to be ineluctable. Since it is also significant and urgent—after all, we don't know how long it will take to outline a sufficiently complete solution to the control problem—this makes it a Great Challenge on the present conception.[31]

* * *

Before concluding this section, it may be helpful to more abstractly categorize the three Great Challenges within a tripartite scheme that includes: *context risks*, *state risks*, and *step risks*. An example of the former is environmental degradation, the first Great Challenge. The reason is that while this poses a number of direct threats to human well-being, it even more importantly *frames* our existential situation on the planet. In this respect it has the capacity to *modulate* the probability of phenomena like inter-state conflicts, civil wars, terrorist attacks, and so on (Torres 2017a). For this reason, I would argue that context risks may be the most urgent of all the urgent threats to our survival.[32] In contrast, the distribution of unprecedented offensive capabilities across society constitutes a state risk because it arises from being in a particular state or configuration; in this case, the configuration of many actors with the unilateral capacity to inflict civilizational harm. As the calculations of subsection 2.2 illustrate, the longer one is exposed to a state risk, the higher the likelihood of disaster. Lastly, superintelligence appears to constitute a step risk because it is associated with transitioning between two states (Bostrom 2014). If humanity solves the control problem and creates a friendly superintelligence, the danger could very well fall to zero—indeed,

---

[30] Other surveys with similar results include Baum et al. 2011 and Sandberg and Bostrom 2011.

[31] For an excellent, comprehensive, and authoritative overview of this topic, see Sotala and Yampolskiy 2015.

[32] More recently, Seán Ó hÉigeartaigh (2017) has used the term "stressor" to describe this category of phenomena. In his words, "we might also consider less severe climate change as a stressor, as it could be expected to lead to major droughts and famines and other resource shortages, mass migration, geopolitical tension that could result in local or global war, and so forth. It could also lead to international conflict, for example over the use of controversial mitigation techniques such as sulphate aerosol geoengineering technologies."

there are reasons for expecting a post-singularity world to be "utopian." Understanding these three distinctions could influence our strategies for neutralizing the corresponding threats, e.g., by compelling us to further prioritize one over the others.

3. Important Neglected Additional Issues

It is part of the excessive egotism of the human animal that the bare idea of its extinction seems incredible to it.—H.G. Wells

This paper has so far explored the three existential risk phenomena that satisfy the tripartite criteria of the Great Challenges framework. But the issues that are germane to understanding our existential predicament—and thus to mitigating those events that could forever preclude the realization of astronomical amounts of value—are far more complex than this already-quite-complex picture may suggest. In this section, I want to examine a number of complicating factors that (a) are directly relevant to overcoming the Great Challenges, and (b) have not been adequately discussed, in my judgment, by scholars focused on the long-term future of humanity. Thus, I will ignore issues such as, e.g., the cognitive biases that can distort perceptions of existential risk; the special difficulty of motivating existential risk mitigation given that doing so yields an intergenerational global public good; and so on. (For books that cover such topics, see Bostrom and Cirkovic 2008; Häggström 2016; Torres 2017a.) This being said, let's begin by briefly switching perspectives from the *a posteriori* to the *a priori*.

(i) *Doomsday Arguments*. The considerations above paint a rather dismal picture of our collective future. But the situation may be even more ominous. Consider the two most prominent theories about self-locating beliefs, namely, the "self-sampling assumption" (SSA) and the "self-indication assumption" (SIA). The former—as well as its more sophisticated sibling, the "strong self-sampling assumption" (SSSA)—yields the Carter-Leslie Doomsday Argument, which concludes that we are systematically underestimating the probability of doom. That is, whatever our empirical estimates of annihilation are, we ought to increase them by some amount (see Leslie 1996).[33] This unwelcome conclusion can be avoided by SIA, but at an even greater cost, since when one combines SIA with the Great Filter framework, SIA implies that *the end is nigh* (Grace 2010; Hanson 2010).[34] For the present purposes, I won't elaborate the philosophical esoterica behind these assertions; suffice it to say that whether one adopts SSA, SSSA, or SIA, the prospects for human survival appear even dimmer than empirical analyses alone might imply. This is a point that any comprehensive examination of our collective survival plight ought to take seriously.

(ii) *Environmentally-Mediated Intellectual Decline*. Returning to the empirical, there are several sequelae of the first Great Challenge enumerated in subsection 2.1 that could nontrivially complicate efforts to mitigate all three Great Challenges. For example, preliminary evidence suggests that the carbon emissions of civilization are not only causing catastrophic climatic and ecological changes, but quite literally making us "dumber" (Grossman 2016). One study, for example, found "moderate" declines in cognitive performance on decision tasks when the ambient concentration of CO2 increased from 600 to 1,000 ppm, and an "astonishingly large" drop in performance from 1,000 to 2,500 ppm (Grossman 2016; see also Romm 2014; Satish et al. 2012; Allen et al. 2016; Torres 2017a). By comparison, our ancestors spent ~200,000 years breathing air with between 180 and 280 ppm of CO2. As mentioned in subsection 2.1, the planet recently passed the ominous milestone of 400 ppm of CO2 in the air, which is irreversible in the foreseeable future, and CO2 levels could reach upwards of 1,000 ppm by the end of this century (IPCC 2018; Torres 2017a). It follows that *there could be widespread, nontrivial deficits in our capacities to solve problems at precisely the moment when we will be forced to confront issues of unprecedented magnitude and complexity* that could irreversibly compromise our future potential.

The situation is even worse than this, though, for the squishy 3-pound computers between our ears. In the last 40 years or so, more than 20,000 new chemicals have been introduced to the market without being tested for possible toxic, including neurotoxic, properties. The troubling fact is that contempo-

---

[33] See Häggström 2016, chapter 7, for a cogent critique of the Doomsday Argument.

[34] For a rebuttal of this conclusion, see Armstrong 2010.

rary humans are exposed to a staggering number of molecules that our ancestors even a few centuries ago would never have encountered. David Bellinger even calculates that "Americans have collectively forfeited" 41 million IQ points "as a result of exposure to lead, mercury, and organophosphate pesticides" alone (see Hamblin 2014). Other chemicals that are known to threaten the brain include arsenic, toluene, DDT/DDE, tetrachloroethylene, cadmium, PBDEs, methanol, ethanol, acrylamide, chlorpyrifos,[35] manganese, PCBs, BPA, fluoride, and, perhaps, some of the prescription drugs, including anti-psychotics, that can be found in public drinking water (see Boerner 2014). Some of these are quite common in our contemporary milieu, including BPA in thermal paper receipts, PCBs in high-fat foods, and fluoride in tap water. Even more, studies have linked a large number of common phenomena to cognitive impairment, including highway pollution, junk food, artificial baby food, nutrient deficiency, excess dietary glucose or fructose, mental illnesses like anxiety and depression, chronic stress, chronic insomnia, and chronic jet lag.[36] The result of exposure to one or more of these phenomena could be what Christopher Williams calls "environmentally-mediated intellectual decline" (EMID). This has both *positive* and *negative* manifestations: The former occurs when, e.g., one is exposed to heavy metals; the latter occurs when, e.g., one suffers from malnutrition (Williams 1997). Sadly, denizens of the developing world are far more susceptible to EMID than those in the developed world, although individuals in both the developing and developed world must still contend with the deleterious cognitive effects of some combination of these phenomena.

On the societal level, even sub-clinical losses of IQ could have significant, wide-ranging effects. As Bostrom (2008) observes, a nootropic drug that improves cognitive performance by only 1 percent "would hardly be noticeable in a single individual," yet "if the 10 million scientists in the world all benefited from the drug the inventor [of the drug] would increase the rate of scientific progress by roughly the same amount as adding 100,000 new scientists." It follows that *subtracting* 1 percent of individual performance would be equivalent to *removing* 100,000 scientists—a potentially very bad outcome with respect to our collective capacity to overcome the Great Challenges.

(iii) *Temporal Discounting*. Yet another under-discussed problem associated with the first Great Challenge is that, according to several studies, people discount the future more as environmental instability increases and life expectancy falls (see, e.g., Daly and Wilson 2005). As Pinker (2011) puts the point, "it doesn't pay to save for tomorrow if tomorrow will never come, or if your world is so chaotic that you have no confidence you would get your savings back." This suggests that the context risk of environmental degradation could *decrease interest* in the general topic of "long-termism," which subsumes issues relating to existential risks, as societies become increasingly preoccupied with more immediate survival concerns. In other words, the correlation between environmental instability and steeper discount rates could be bad news for existential risk research and, therefore, bad news for the long-term perpetuation of our lineage.

The worrisomeness of this situation is further underlined by research that reveals that existential risks are already a severely neglected topic that few people are actively studying. As Bostrom (2013) notes, there were, as of 2012, far more scholarly articles about dung beetles than about human extinction. Along these lines, I conducted some Google Scholar searches and found that, as of January 24, 2018, there were exactly 1,910 results for the word "existential risk." In comparison, there were 2,060 results for "Super Mario Brothers," 2,100 for "dog flea," 2,320 for "French cheese," 8,760 for "anal penetration," 12,800 for "FOXP2," 66,800 for "bicuspid," and 170,000 for "hospitality management"—all of which are dwarfed by the 5,390,000 results for "cancer." While cancer research is obviously important for human well-being, much of its value is predicated on the continued existence of humanity, since most of its benefits won't be fully realized for many generations. This implies that ensuring human survival should take precedence over curing cancer—yet just the opposite is the case.

---

[35] Sadly, the Trump administration recently reversed an EPA ban on chlorpyrifos—a kind of organophosphate that likely has negative neuro-developmental effects—after Dow Chemicals, a manufacturer of the chemical, donated 1 million dollars to Trump's inauguration fund (see Snopes 2017).

[36] Note also that some studies suggest that human intelligence has been in decline for millennia. As Gerald Crabtree (2013) writes, "I would be willing to wager that if an average citizen from Athens of 1000 BC were to appear suddenly among us, he or she would be among the brightest and most intellectually alive of our colleagues and companions."

(iv) *Intractable Complexity*. A problem that is pertinent to all three Great Challenges concerns the rapid *complexification* of the contemporary world since the Industrial Revolution and especially, according to some scholars, the 1980s. First, consider the general epistemic situation of contemporary humans. As one author observes,

> it was possible as recently as three hundred years ago for one highly learned individual to know everything worth knowing. By the 1940s, it was possible for an individual to know an entire field, such as psychology. Today the knowledge explosion makes it impossible for one person to master even a significant fraction of one small area of one discipline (Jacobs 2003).

The implications of this observation are profound. Consider the following line of reasoning: (a) The term "knowledge explosion" above refers to the exponential growth of *collective human knowledge*; (b) individual humans are subject to what Christopher Cherniak (1992) calls our "finitary predicament," which results from hard limits on the temporal and cognitive resources that are available to us;[37] (c) it follows from (b) that the capacity for *individual human knowledge* has not grown exponentially over time;[38] (d) the phenomenon of (a) thus entails a corresponding "ignorance explosion," where this term refers to the exponential divergence between what is *known* by the collective (at some moment) and what is *knowable* by the individual, given the fact of (c). Robert Oppenheimer makes this point even more explicitly in a 1956 article about the potential power and limits of human reason; to quote him at length:

> Positive knowledge, what is recorded in the technical books and learned journals, all of it that is new and true and not trivial, is of course not wisdom; it can on occasion almost appear incompatible with wisdom. I think that such positive knowledge doubles in less than a generation, perhaps in a decade. This means that most of what there is to know about the world of nature was not discovered when a man went to school; it means that universal knowledge, always, even in Leonardo's day, a dream, but not an irrelevant dream, has become a mockery; it means that the specialized sciences, genetics, for instance, or astrophysics, or mathematics, are like the fingers of a hand: they all arise from the common matrix of common sense, from man's daily experience, his history, his tradition, and his words. Each is now developing a life, an experience, and a language of its own, and between the tips of the fingers there is rare contact. For many centuries mathematics and physics grew in each other's company, in happy symbiosis. Today at their growing tips they hardly touch. Logic, psychology, philosophy were long studied in the same rooms, and often by the same [person]. Today they rarely speak to each other, and are more rarely understood or even heard. The deep, detailed, intimate almost loving knowledge of a specialized science is lost in synoptic views of science as a whole. These changes mean that ignorance is a universal, pervasive feature of our time. It is clear that they have an essential relevance to the problems of education (Oppenheimer 1956).

This—the ignorance explosion, resulting in "universal, pervasive" *nescience*—matters because competence at the individual level requires knowledge at the individual level, yet knowledge at the individual level is shrinking as a proportion of total knowledge. Hence, individuals are becoming less competent over time. The exact same could be said of *groups* that constitute subsets of the collective whole: Even an organization consisting of many capable intellectuals will be incapable of meaningfully grasping more than a small fraction of what is currently known about the universe (and all it envelopes), and this will seriously undercut its ability to make competent decisions. This is, generally speaking, a bad situation for humanity to find itself in, since it suggests that, while everyone paddles, no one is actually steering the ship. Indeed, no one *could*.

With respect to existential risk studies in particular, the consequences of the ignorance explosion are especially acute. The reason is that the more interdisciplinary a field, the greater the impact of individual ignorance, given the "epistemic breadth-depth tradeoff" (i.e., crudely put, one can know a lot about

---

[37] Perhaps the Flynn effect constitutes a small exception.

[38] A fact that could change with the advent of cognitive enhancements. See below.

a little, or a little about a lot). Indeed, a central aim of this nascent but important field is to determine which directions humanity should steer the ship, insofar as we have control over our trajectory through time. But acquiring robust knowledge about high-level, big-picture issues that span so many fields of human inquiry—from economics to biology, astrobiology to population ethics, computer science to sociology, technology studies to decision theory, and so on[39]—is, as Jacobs implies above, increasingly beyond the bounds of human capability. In a phrase, the complexification of the human condition on this pale blue dot is reducing the capacity for existential risk research to provide precisely the sort of insights needed to ensure a good outcome for our lineage.[40]

Yaneer Bar-Yam (2002) makes a similar point in the specific context of human governance. He argues that societies have, partly because of specialization since the 1980s, transformed from hierarchically-controlled systems to networked systems, where a hallmark of the latter is lateral (rather than vertical) interactions between individuals and "subgroups." The result has been a staggering increase in the overall complexity of society to the point that no single individual or group can effectively control, regulate, direct, or coordinate collective behaviors. In Bar-Yam's (2002) words,

> complex systems that display complex collective behavior are structured as networks. By contrast, the traditional human social structure, whether in government or in industry, has been based upon control hierarchies. Just as a single neuron is not able to dictate the behavior of a neural system, an emergent complex network of human beings may not be directed by a single human being.

In other words, human civilization is becoming *ungovernable* by *anyone*, however individually "competent" she or he might be. This has implications for what Tegmark, following Sagan, refers to as "the race between the growing power of technology and the growing wisdom with which we manage it," where this race "will determine the fact of humanity" (Tegmark 2016). If Bar-Yam is correct, then the race appears to have already been lost—at least by us *humans*.[41] Put in these terms, complexification, which is driven partly by technological innovation, is making it impossible for any individual, however *sage* she or he might be, to make truly wise decisions about how civilization should navigate the obstacle course of existential hazards before us. I am not here claiming that this is the final word on the matter. Rather, I want to point out that this is an urgent issue that deserves far more thoughtful consideration by existential risk scholars than it has thus far received.

(v) *Catastrophe Clustering*. While many analyses of global catastrophic risks (GCRs), of which existential risks are a special type, focus on single-disaster scenarios, there are reasons for expecting catastrophes to temporally cluster together, a phenomenon that I call "*catastrophe clustering*." There are at least five notable reasons for this, some obvious and others not so obvious. The first reason concerns the fact that random events tend to cluster together in time, yielding what psychologists call the "clustering illusion." The question is thus: What reason is there for expecting global catastrophes to be random (meaning that the relevant data set is indistinguishable from a Poisson process)? On the one hand, the timing of many natural phenomena, such as asteroid impacts, appears to be random in some nontrivial sense. If this is true, then the probability of an asteroid B striking Earth the day after an asteroid A struck Earth is higher than the probability of B striking Earth a week later. On the other hand, studies show that both the onset and termination of wars throughout history are randomly timed (see Richardson 1960; Pinker 2011). If this is the case, then we have a *prima facie* reason for expecting anthropogenic catastrophes to be randomly timed as well, thus resulting in catastrophe clusters.

---

[39] Or, as I sometimes more succinctly put it, *from quantum theory to quantal theory* (in my home field of neuroscience).

[40] Perhaps, then, one of the main priorities of existential risk mitigation efforts should be the development of safe and effective cognitive enhancements for the purpose of better approximating individual and collective knowledge (see Walker 2002; Verdoux 2011; Torres 2017a). A similar benefit could be obtained through the creation of a friendly superintelligence, as alluded to in subsection 2.3.

[41] That is to say, there could be cognitively enhanced posthumans, advanced algocratic systems, or even a "friendly super-singleton" that overcomes the complexity problems here discussed; see Torres 2017d.

Second, an issue that pertains to aforementioned issue of complexification: The unprecedented, and quickly growing, interconnectedness of the contemporary world could produce cascading effects between coupled networks, resulting in "network of networks" disasters—the most extreme being "Black Swans" (Taleb 2007)—that ripple throughout the entire global system. As Shimizu and Clark (2015) observe,

> the decadal trends and the best available science all clearly indicate that geophysical, meteorological, biological, technological, and human induced disasters are increasing in intensity (also many in frequency), complexity (interconnected, synergistic, and cascading), [and] uncertainty (future new events). Further, these multiplying risk factors are interacting with an ever more complex set of physical, social, economic and environmental vulnerabilities at rates that nations, societies, and commerce are ill-prepared to deal with in terms of "gaps" in existing governance and institutional capacities.

The 2005 Katrina disaster in the US and the 2011 Tohoku disaster in Japan provide paradigm cases of how devastating cascading disasters can be (Shimizu and Clark 2015). As globalization further increases the interconnectedness of different networks, the danger that single component failures could amplify through the larger system may increase.

Third, a non-existential disaster that is nonetheless catastrophic could temporarily compromise the resiliency of society, thus rendering it unusually vulnerable to a second disaster. In this case, the second disaster might not have otherwise been catastrophic; it is the loss of preventative mechanisms that (negatively) causes the majority of damage. For example, a global crop failure could lead to widespread malnutrition, and this could facilitate the spread of infectious diseases, thus turning a local epidemic into a global pandemic.[42] In fact, numerous studies suggest that climate change could cause multiple devastating, simultaneous crop failures around the world, which would cause global food supply disruptions (see, e.g., Tigchelaar 2018; Scheelbeek et al. 2018). Meanwhile, "public health experts believe we are at greater risk than ever of experiencing large-scale outbreaks and global pandemics like those we've seen before: SARS, swine flu, Ebola, and Zika" (Senthilingam 2017). Thus, not only could the former increase the probability of the latter, but there could emerge synergistic effects between the two that kill more people than each would kill separately, added together.

Fourth, a point about agential risks: if a large-scale catastrophe increases the vulnerability of society to subsequent, temporally-proximate disasters, then the occurrence of a large-scale catastrophe could *trigger* terror agents to opportunistically launch an attack. Put somewhat simplistically, intentionally destroying the world *entirely* may be much easier if the world has already been destroyed *partially*. One should thus be especially vigilant of shrewd agents who are motivated by omnicidal or anti-civilizational ideologies if a global catastrophe were to happen in the future. In fact, studies show that terrorist attacks have historically often triggered subsequent attacks (Jenkins et al. 2016), and mass killings are known to increase the probability of copycat shootings for 13 days on average after the initial event (Towers et al. 2015).

And fifth, a more scenario-specific possibility: Seth Baum and colleagues have outlined a hypothetical "double catastrophe scenario" in which an ongoing "solar radiation management" (SRM) project is interrupted by a destabilizing event—e.g., a terrorist attack, interstate or civil war, shift of political power, economic recession, and so on. Suddenly terminating a stratospheric geoengineering project could wreak unpredictable havoc on the global climate, bringing about massive agricultural failures or, at the extreme, initiating a runaway greenhouse effect (Baum et al. 2013).

Not only are some of these causal models plausible, but the redundancy of potential causes for catastrophe clustering suggests that humanity should brace itself for shocks following a global disaster, whether natural or anthropogenic in origin.

(vi) *Men*. Another issue that has not been adequately addressed in the existential risk literature is *toxic masculinity*. Consider that the overwhelming number of interstate and civil wars, terrorist attacks, rampage shootings, serial killings, homicides, assaults, rapes, instances of domestic violence, acts of cru-

---

[42] Thanks to Karin Kuhlemann for pointing out some of possibilities that I'd overlooked.

elty toward animals, and hate crimes are perpetrated by men. (Not to mention that male psychopaths outnumber female psychopaths by 20 to 1.) The only category of crime that women consistently have higher arrest rates for is prostitution. Such facts lead Carl Sagan, following Alan Alda, to warn about the immense dangers of "testosterone poisoning," which can cause abnormally intense aggression and violence. Similarly, David Pearce (2012) eloquently writes that

> the single greatest underlying risk to the future of intelligent life isn't technological, but both natural and evolutionarily ancient, namely competitive male [dominance] behaviour. Crudely speaking, evolution "designed" human male primates to be hunters/warriors. Adult male humans are still endowed with the hunter-warrior biology—and primitive psychology—of our hominin ancestors. For the foreseeable future, all technological threats must be viewed through this sinister lens. Last century, male humans killed over 100 million fellow humans in conflict and billions of non-humans. Directly or indirectly, this century we are likely to kill many more. But perhaps we'll do so in more sophisticated ways.

Along these lines, Persson and Savulescu argue that moral bioenhancement interventions should target men in particular, given their statistically significant violent proclivities and moral deficits. As they write, "if it is right that women are more altruistic than men, it seems that we could make men in general more moral by making them more like women by biomedical methods, or rather, more like the men who are more like women in respect of empathy and aggression" (Persson and Savulescu 2011). I have elsewhere criticized Persson and Savulescu's moral bioenhancement thesis, but these criticisms do not concern their point about men and women, the latter of whom are consistently underrepresented in decisions to start wars and make peace and, indeed, should play a much larger role in shaping the developmental trajectory of civilization. Even more, in terms of overcoming the complexity of studying the future of humanity, a 2010 study found that groups can exhibit a single, measurable property that is exactly analogous to psychometric *g* in individuals. Yet the *only* member-level variable that this study found to be directly and positively correlated with group-level "collective intelligence" is the number of women within the group. That is to say, the more women, the smarter the group (Woolley et al. 2010). It follows that insofar as existential risk mitigation is a group activity, the research community should strive to include more women.

In sum, a male-dominated geopolitical arena in which unprecedentedly powerful technologies are increasingly accessible to state and nonstate actors could be exceptionally susceptible to doom. Toxic masculinity, testosterone poisoning, and the relative statistical lack of empathy/sympathetic concern among men pose serious threats to human survival. There is, indeed, an urgent need for trenchant feminist critiques of existential risk studies in general, a task that the present author is, unfortunately, ill-equipped to accomplish.

(vii) *Space Colonization*. Finally, I should add here that one of the primary sources of existential hope for many future-oriented thinkers—including Elon Musk, Stephen Hawking, Derek Parfit, Nick Bostrom, Carl Sagan, Richard Gott, Jason Matheny, and Lord Martin Rees—is the belief that spreading into space will significantly reduce the probability of doom before our lineage crosses certain thermodynamic thresholds of unlivability *far* in the future (say, 10^{40} years; see Adams 2008). From an evolutionary biology perspective, this strategy makes sense. But in a recent paper, which draws from ideas first explored by Daniel Deudney in *Dark Skies* (forthcoming), I outline a number of theoretically strong reasons for suspecting that space colonization will almost certainly result in a suffering catastrophe, if not the total annihilation of all life in the universe (Torres 2018; Deudney, forthcoming). I won't delineate this argument here, but suffice it to say that expanding into the heavens may not constitute the "existential panacea" that many space expansionists believe that it is. Note that if I am correct about this, then the "astronomical trajectories, in which human civilization expands beyond its home planet and into the accessible portions of the cosmos," as discussed by Baum et al. in this issue of *Foresight*, constitute a likely subtype of "catastrophe trajectories, in which one or more events cause significant harm to human civilization" (Baum et al. 2018). Humanity is ultimately doomed if we remain on Earth, but perhaps this fate is actually preferable to one that realizes a "suffering risk," which would entail the realization of astronomical amounts of suffering (see Althaus and Gloor 2018; Tomasik 2017)—or, even worse, a "hyperexistential catastrophe," whereby extinction would be unambiguously better than survival. Both of these are real

outcomes if humanity ventures into and beyond the solar system, the Milky Way galaxy, the Virgo Supercluster, and so on.

4. Conclusion

This paper aims to offer an authoritative overview of the growing obstacle course of dangers before us. It argues that three broad classes of threats satisfy the criteria of being significant, urgent, and ineluctable, and that humanity should prioritize these challenges over other problems—that is, insofar as one accepts the maxipok rule. The paper then examined a number of complicating factors that, in my view, are both (a) important to have in the foreground of futurological thinking, and (b) have not received much attention in the scholarly literature. In conclusion, there is hardly a single problem facing humanity today that is fundamentally insoluble. The question is whether humanity will muster the wisdom and foresight to ensure future human flourishing.

References:

Adams, Fred. 2008. Long-Term Astrophysical Processes. In Nick Bostrom and Milan Cirkovic (eds.), *Global Catastrophic Risks*. Oxford: Oxford University Press.

Allen, Joseph, Piers MacNaughton, Usha Satish, Suresh Santanam, Jose Vallarino, and John Spengler. 2016. Associations of Cognitive Function Scores with Carbon Dioxide, Ventilation, and Volatile Organic Compound Exposures in Office Workers: A Controlled Exposure Study of Green and Conventional Office Environments. *Environmental Health Perspectives*. 124(6): 805-812.

Althaus, David, and Lukas Gloor. 2018. Reducing Risks of Astronomical Suffering: A Neglected Priority. Foundational Research Institute. https://foundational-research.org/reducing-risks-of-astronomical-suffering-a-neglected-priority/.

Alvaredo, Facundo, Lucas Chancel, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. 2018. World Inequality Report 2018. http://wir2018.wid.world/files/download/wir2018-full-report-english.pdf.

Armstrong, Stuart. 2010. SIA Won't Doom You. *LessWrong*. https://www.lesswrong.com/posts/vaZYAs7tsDriNkoAP/sia-won-t-doom-you.

Armstrong, Stuart, and Kaj Sotala. 2012. How We're Predicting AI—or Failing To. Machine Intelligence Research Institute. https://pdfs.semanticscholar.org/5a12/80f783e4ce6ba31b821f4d86f612ef733213.pdf.

Barkham, Patrick. 2018. "We're doomed": Mayer Hillman on the climate reality no one else will dare mention. *Guardian*. https://www.theguardian.com/environment/2018/apr/26/were-doomed-mayer-hillman-on-the-climate-reality-no-one-else-will-dare-mention?CMP=share_btn_fb.

Barnosky, A.D., Sadly, E.A., Bascompte, J., Berlow, E.L., Brown, J.H., Forelius, M., Getz, W.M., Harte, J., Hastings, A., Marquet, P.A., Martinez, N.D., Mooers, A., Roopnarine, P., Vermeij, G., Williams, J.W., Gillespie, R., Kitzes, J., Marshall, C., Matzke, N., Minder, D.P., Revilla, E., Smith, A.B., 2012 Approaching a State Shift in Earth's Biosphere, Nature 486 52-58.

Baum, Seth, and Ben Goertzel, and Ted Goertzel. 2011. How Long Until Human-Level AI? Results from an Expert Assessment. *Technological Forecasting and Social Change*. 78(1): 185-195.

Baum, Seth, Timothy Maher, and Jacob Haqq-Misra. 2013. Double Catastrophe: Intermittent Stratospheric Geoengineering Induced By Societal Collapse. *Environment, Systems and Decisions*. 33(1): 168-180.

Baum, Seth, Stuart Armstrong, Timoteus Ekenstedt, Olle Häggström, Robin Hanson,

Karin Kuhlemann, Matthijs M. Maas, James D. Miller, Markus Salmela, Anders Sandberg, Kaj Sotala, Phil Torres, Alexey Turchin, and Roman V. Yampolskiy. 2018. Long-Term Trajectories of Human Civilization. Forthcoming in *Foresight*.

Becatoros, E., 2017 More than 90 Percent of World's Coral Reefs Will Die by 2050, *Independent*. http://www.independent.co.uk/environment/environment-90-percent-coral-reefs-die-2050-climate-change-bleaching-pollution-a7626911.html.

Beckstead, Nick. 2013. On the Overwhelming Importance of Shaping the Far Future. Dissertation. https://rucore.libraries.rutgers.edu/rutgers-lib/40469/PDF/1/play/.

Boerner, Leigh Krietsch. 2014. The Complicated Question of Drugs in the Water. *Nova Next*. http://www.pbs.org/wgbh/nova/next/body/pharmaceuticals-in-the-water/.

Böhm, Monika, Ben Collen, Jonathan Baillie, Philip Bowles, Janice Chanson, Neil Cox, Geoffrey Hammerson, Michael Hoffmann, Suzanne Livingstone, Mala Ram, Anders Rhodin, Simon Stuart, Peter Paul van Dijk, Bruce Young, Leticia Afuang, Aram Aghasyan, Andrés García, César Aguilar, … George Zugcy. 2013. The Conservation Status of the World's Reptiles. *Biological Conservation*. 157: 372-385.

Bostrom, Nick. 2002 Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards, J. Evo. Tech. (9)1.

Bostrom, N. 2003a Are You Living in a Computer Simulation?, Philo. Quarterly 53(211): 243-255.

Bostrom, Nick. 2005. A Philosophical Quest for Our Biggest Problems. TED. https://www.ted.com/talks/nick_bostrom_on_our_biggest_problems/transcript.

Bostrom, Nick. 2008. Three Ways to Advance Science. *Nature*. https://nickbostrom.com/views/science.pdf.

Bostrom, N. 2013 Existential Risk Prevention as Global Priority, Glo. Pol. 4(1): 15-31.

Bostrom, N. 2014 *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Bostrom, Nick, and Milan Ćirković. 2008. *Global Catastrophic Risks*. Oxford, UK: Oxford University Press.

Bourget, David, and David Chalmers. 2014. What Do Philosophers Believe? *Philosophical Studies*. 170(3): 465-500.

Carlson, R. 2014 Time for New DNA Synthesis and Sequencing Cost Curves, Synbiobeta. https://synbiobeta.com/time-new-dna-synthesis-sequencing-cost-curves-rob-carlson/.

Carrington, Damian. 2018. Paul Ehrlich: "Collapse of Civilization is a Near Certainty Within Decades." *Guardian*. https://www.theguardian.com/cities/2018/mar/22/collapse-civilisation-near-certain-decades-population-bomb-paul-ehrlich.

Ceballos, Gerardo, Paul Ehrlich, Anthony Barnosky, Andrés García, Robert Pringle, and Todd M. Palmer. 2015. Accelerated Modern Human-Induced Species Losses: Entering the Sixth Mass Extinction. *Science Advances*. 1(5).

Center. 2018. The Extinction Crisis. Center for Biological Diversity. http://www.biologicaldiversity.org/programs/biodiversity/elements_of_biodiversity/extinction_crisis/.

Chalmers, D. 2010 The Singularity: A Philosophical Analysis, J. Consc. Studies 17(9-10): 7-65.

Cherniak, Christopher. 1992. *Minimal Rationality*. Cambridge, MA: The MIT Press.

Cooper, J. 2013 Bioterrorism and the Fermi Paradox, *International Journal of Astrobiology*. 12(2): 144-148.

Crabtree, Gerald. 2013. Our Fragile Intellect. Part 1. *Trends in Genetics*. 29(1): 1-3.

Daly, Martin, and Margo Wilson. 2005. Carpe Diem: Adaptation and Devaluing the Future. *The Quarterly Review of Biology*. 80(1): 55-61.

Deudney, D. forthcoming. *Dark Skies: Space Expansionism, Planetary Geopolitics, and the End of Humanity*, Oxford University Press.

Diamandis, Peter. 2014. *Abundance: The Future is Better Than You Think*. New York, NY: Free Press.

Diamond, J. 2005 Collapse: How Societies Choose to Fail or Succeed, Viking.

DoW. 2018. Coral Reef Fish. Defenders of Wildlife. https://defenders.org/coral-reef/coral-reef-fish.

Dye, LaVonne. 1993. The Marine Mammal Protection Act: Maintaining the Commitment to Marine Mammal Conservation. *Case Western Reserve Law Review*. 43(4): 1411-1448.

Edwards, Lin. 2010. Humans Will Be Extinction in 100 Years Says Eminent Scientist. *Physorg*. https://phys.org/news/2010-06-humans-extinct-years-eminent-scientist.html.

Ehgartner, Ulrike, Patrick Gould, and Marc Hudson. 2017. On the Obsolescence of Human Beings in Sustainable Development. *Global Discourse*. 7(1): 66-83.

Estrada, Alejandro, Paul Garber, Anthony Rylands, Christian Roos, Eduardo Fernandez-Duque, Anthony Di Fiore, K. Anne-Isola Nekaris, Vincent Nijman, Eckhard Heymann, Joanna Lambert, Francesco Rovero, Claudia Barelli, Joanna Setchell, Thomas R. Gillespie, Russell Mittermeier, Luis Verde Arregoitia, Miguel de Guinea, Sidney Gouveia, Ricardo Dobrovolski, Sam Shanee, Noga Shanee, Sarah Boyle, Agustin Fuentes, Katherine C. MacKinnon0 Katherine Amato, Andreas Meyer, Serge Wich, Robert Sussman, Ruliang Pan, Inza Kone, and Baoguo Li. 2017. Impending Extinction Crisis of the World's Primates: Why Primates Matter. *Science Advances*. 3(1).

Ehrlich, Paul, and Anne Ehrlich. 2009. The Population Bomb Revisited. *Electronic Journal of Sustainable Development*. 1(3): 63-71.

Etzioni, Oren. 2016. No, the Experts Don't Think Superintelligent AI is a Threat to Humanity. *MIT Technology Review*. https://www.technologyreview.com/s/602410/no-the-experts-dont-think-superintelligent-ai-is-a-threat-to-humanity/.

Farquhar, Sebastian, John Halstead, Owen Cotton-Barratt, Stefan Schubert, Haydn Belfield, and Andrew Snyder-Beattie. 2017. Existential Risk: Diplomacy and Governance. Global Priorities Project. https://www.fhi.ox.ac.uk/wp-content/uploads/Existential-Risks-2017-01-23.pdf.

Fecht, Sarah. 2017. Stephen Hawking Says We Have 100 Years to Colonize a New Planet—Or Die. Could We Do It? *Popular Science*. https://www.popsci.com/stephen-hawking-human-extinction-colonize-mars.

FOP. 2015. Learning to Die in the Anthropocene: Interview with Roy Scranton. Friends of the Pleistocene. https://fopnews.wordpress.com/2015/09/24/scrantonanthropocene/.

Frick, Johann. 2017. On the Survival of Humanity. *Canadian Journal of Philosophy*. 47(2-3): 344-367.

Gallucci, Robert. 2005. Averting Nuclear Catastrophe. *Harvard International Review*. http://hir.harvard.edu/article/?a=1303.

Garvey, Megan. 2014. Transcript of the Disturbing Video "Elliot Roger's Retribution." *LA Times*. http://www.latimes.com/local/lanow/la-me-ln-transcript-ucsb-shootings-video-20140524-story.html.

GBO-3. 2010. Global Biodiversity Outlook 3. URL: https://www.cbd.int/doc/publications/gbo/gbo3-nal-en.pdf.

Gohd, Chelsea. 2017. Elon Musk Claims We Only Have a 10 Percent Chance of Making AI Safe. *Futurism*. https://futurism.com/elon-musk-claims-only-have-10-percent-chance-making-ai-safe/.

Grace, Katja. 2010 SIA Doomsday: The Filter is Ahead, Meteuphoric. https://meteuphoric.wordpress.com/2010/03/23/sia-doomsday-the-filter-is-ahead/.

Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. 2017. When Will AI Exceed Human Performance? Evidence from AI Experts. *arXiv*. https://arxiv.org/pdf/1705.08807.pdf.

Graham, Bob, Jim Talent, Graham Allison, Robin Cleveland, Steve Rademaker, Tim Roemer, Wendy Shewrman, Henry Sokolski, and Rich Verma. 2008. World at Risk: The Report of the Commission on the Prevention of WMD Proliferation and Terrorism. http://www.dtic.mil/dtic/tr/fulltext/u2/a510559.pdf.

Grossman, Daniel. 2016. High CO2 Levels Inside and Out: Double Whammy? *Yale Climate Connections*. https://www.yaleclimateconnections.org/2016/07/indoor-co2-dumb-and-dumber/.

Häggström, Olle. 2016. *Here Be Dragons: Science, Technology and the Future of Humanity*. Oxford, UK: Oxford University Press.

Hamblin, James. 2014. The Toxins that Threaten Our Brains. *The Atlantic*. https://www.theatlantic.com/health/archive/2014/03/the-toxins-that-threaten-our-brains/284466/.

Hand, Eric. 2016. Could Bright, Foamy Wakes from Ocean Ships Combat Global Warming? *Science*. URL: http://www.sciencemag.org/ news/2016/01/could-bright-foamy-wakes-ocean-ships-combat-global-warming.

Hanson, R. 2010 Very Bad News, Overcoming Bias. http://www.overcomingbias.com/2010/03/very-bad-news.html.

Haqq-Misra, Jacob, Sanjoy Som, Brendan Mullan, Rafael Loureiro, Edward Schwieterman, Lauren Seyler, and Haritina Mogosanu. 2018. The Astrobiology of the Anthropocene. https://arxiv.org/pdf/1801.00052.pdf.

Harris, Sam. 2018. What Is and What Matters. *Waking Up*. https://samharris.org/podcasts/108702/.

Hawking, S. 2016 This is the Most Dangerous Time for Our Planet, Guardian. https://www.theguardian.com/commentisfree/2016/dec/01/stephen-hawking-dangerous-time-planet-inequality.

Hersher, R. 2016 Elon Musk Unveils His Plan for Colonizing Mars, NPR. http://www.npr.org/sections/thetwo-way/2016/09/27/495622695/this-afternoon-elon-musk-unveils-his-plan-for-colonizing-mars.

IPCC. 2014 Climate Change 2014 Synthesis Report. https://www.ipcc.ch/news_and_events/docs/ar5/ar5_syr_headlines_en.pdf.

IPCC. 2018. Carbon Dioxide: Projected Emissions and Concentrations. Accessed on 1/27/2018. http://www.ipcc-data.org/observ/ddc_co2.html.

Jacobs, Gregg. 2003. *The Ancestral Mind*. London, UK: Penguin.

Jacquet, Jennifer. 2017. The Anthropocene. *The Edge*. https://www.edge.org/response-detail/27096.

Jamail, Dahr. 2013. Tomgram: Dahr Jamail, The Climate Change Scorecard. *TomDispatch*. http://www.-tomdispatch.com/blog/175785/tomgram%3A_dahr_jamail%2C_the_climate_change_scorecard/.

Jenkins, Brian Michael, Henry Willis, and Bing Han. 2016. Do Significant Terrorist Attacks Increase the Risk of Further Attacks? Initial Observations from a Statistical Analysis of Terrorist Attacks in the United States and Europe from 1970 to 2013. RAND. https://www.rand.org/content/dam/rand/pubs/perspectives/PE100/PE173/RAND_PE173.pdf.

Kelly, Mark. 2017. This Year Has Been an Unequivocal Disaster for the Future of the Planet. CNN. http://www.cnn.com/2017/12/26/opinions/earth-from-space-climate-change-opinion-mark-kelly/index.html.

Kuhlemann, Karin. 2018. We Can't Tackle Overpopulation when the Time Comes—We Need to Talk About it Now. *Huffington Post*. http://www.huffingtonpost.co.uk/entry/lets-stop-thinking-we-can-tackle-it-when-the-time-comes-we-need-to-talk-about-overpopulation-now_uk_5a675db0e4b002283006fe0c.

Kurzweil, Ray. 2005. *The Singularity Is Near*. New York, NY: Penguin Group.

Leslie, J. 1996 The End of the World: The Science and Ethics of Human Extinction, Routledge.

Levitan, D. 2012 After Extensive Mathematical Modeling, Scientist Declares "Earth is F**ked," io9. https://io9.gizmodo.com/5966689/after-extensive-mathematical-modeling-scientist-declares-earth-is-fucked.

Lombroso, P. 2016 Chomsky: "Republicans Are a Danger to the Human Species," il manifesto (Global Edition). https://global.ilmanifesto.it/chomsky-republicans-are-a-danger-to-the-human-species/.

Mayr, Ernst. 1995. Can SETI Succeed? Not Likely. Planetary Society's Bioastronomy News. 7(3).

Mecklin, J. 2018. It is 2 Minutes to Midnight. *Bulletin of the Atomic Scientists*. https://thebulletin.org/sites/default/files/2018%20Doomsday%20Clock%20Statement.pdf.

Miller, Erin. 2015. Mass-Fatality, Coordinated Attacks Worldwide, and Terrorism in France. National Consortium for the Study of Terrorism and Responses to Terrorism. https://www.start.umd.edu/pubs/START_ParisMassCasualtyCoordinatedAttack_Nov2015.pdf.

MIT. 2017. Genetic Engineering Holds the Power to Save Humanity or Kill It, MIT Technology Review. https://www.technologyreview.com/s/608903/genetic-engineering-holds-the-power-to-save-humanity-or-kill-it/.

Mora, Camilo, Bénédicte Dousset, Iain Caldwell, Farrah Powell, Rollan Geronimo, Coral Bielecki, Chelsie Counsell, Bonnie Dietrich, Emily Johnston, Leo Louis, Matthew Lucas, Marie McKenzie, Alessandra Shea, Han Tseng, Thomas Giambelluca, Lisa Leon, Ed Hawkins, and Clay Trauernicht. 2017. Global Risk of Deadly Heat. *Nature*. 7: 501-506.

Mora, Camilo, Randi Rollins, Katie Taladay, Michael Kantar, Mason Chock, Mio Shimada, and Erik Franklin. 2018. Bitcoin Emissions Alone Could Phs Global Warming Above 2°C. *Nature Climate Change*. 8: 931-933.

Motesharrei, S., Rivas, J., Kalnay, E., 2014 Human and Natural Dynamics (HANDY): Modeling Inequality and Use of Resources in the Collapse or Sustainability of Societies, Ecol. Ec. 101: 90-102.

Müller, Vincent, and Nick Bostrom. 2014. Future Progress in Artificial Intelligence: A Survey of Expert Opinion. In Vincent Müller (ed.), *Fundamental Issues of Artificial Intelligence*. Berlin: Springer.

Nuccitelli, Dana. 2018. 2017 Was the Hottest Year on Record Without an El Niño, Thanks to Global Warming. *Guardian*. https://www.theguardian.com/environment/climate-consensus-97-per-cent/2018/jan/02/2017-was-the-hottest-year-on-record-without-an-el-nino-thanks-to-global-warming.

Ó hÉigeartaigh, Seán. 2017. The State of Research in Existential Risk. In John Garrick (ed.), *First International Colloquium on Catastrophic and Existential Risk*.

O'Leary, D., 2006 Escaping the Progress Trap, Geozone Communications.

Oppenheimer, Robert. 1956. Science and Our Times. *Bulletin of the Atomic Scientists*. 12(7): 235-237.

Ord, T. 2015. Will We Cause Our Own Extinction? Natural versus Anthropogenic Extinction Risks. YouTube. https://www.youtube.com/watch?v=uU0Z4psY32s.

Oxfam. 2017. Richest 1 Percent Bagged 82 Percent of Wealth Created Last Year—Poorest Half of Humanity Got Nothing. https://www.oxfam.org/en/pressroom/pressreleases/2018-01-22/richest-1-percent-bagged-82-percent-wealth-created-last-year.

Pamlin, Denis, and Stuart Armstron. 2015. 12 Risks that Threatn Human Civilisation. Global Challenges Foundation. https://api.globalchallenges.org/static/wp-content/ uploads/12-Risks-with-in nite-impact.pdf.

Pantucci, Raffaello. 2011. A Typology of Lone Wolves: Preliminary Analysis of Lone Islamist Terrorists. *Developments in Radicalisation and Political Violence*. http://icsr.info/wp-content/uploads/2012/10/1302002992ICSRPaper_ATypologyofLoneWolves_Pantucci.pdf.

Park, Chang-Eui, Su-Jong Jeong, Manoj Joshi, Timothy Osborn, Chang-Hoi Ho, Shilong Piao, Deliang Chen, Junguo Liu1, Hong Yang, Hoonyoung Park, Baek-Min Kim, and Song Feng. 2017. Keeping Global Warming Within 1.5C Constrains Emergence of Aridification. *Nature Climate Change*. 8: 70-74.

Perry, Tekla. 2018. Move Over Moore's Law, Make Way for Huang's Law. *IEE Spectrum*. https://spectrum.ieee.org/view-from-the-valley/computing/hardware/move-over-moores-law-make-way-for-huangs-law.

Persson, Ingmar, and Julian Savulescu. 2011. Getting Moral Enhancement Right: The Desirability of Moral Bioenhancement. *Bioethics*. 27(3): 124-131.

Persson, I., Savulescu, J. 2012 *Unfit for the Future: The Need for Moral Enhancement*, Oxford University Press, Oxford.

Pinker, Steven. 2011. *The Better Angels of Our Nature: Why Violence Has Declined*. New York, NY: Penguin Books.

Pinker, Steven. 2018. *Enlightenment Now: The Case for Reason, Science, Humanism, and Progress*. New York, NY: Penguin Books.

Posner, R. 2004 Catastrophe: Risks and Response, Oxford University Press.

Potter, Ned. 2009 Can We Grow More Food in 50 Years Than in All of History? ABC News. http://abc-news.go.com/Technology/world-hunger-50-years-food-history/story?id=8736358.

Price, M. 2017. Why Human Society Isn't More—or Less—Violent than in the Past. *Science*. http://www.sciencemag.org/news/2017/12/why-human-society-isn-t-more-or-less-violent-past.

Rawls, John. 1999. *A Theory of Justice*. New York, NY: Belknap Press.

Rees, M. 2003 Our Final Hour: A Scientist's Warning, Basic Books.

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, III, F.S., Lambin, E., Lenton, T.M., Scheffer, M., Folke, C., Schellnhuber, H., Nykvist, B., De Wit, C.A., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P.K., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R.W., Fabry, V. J., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., and Foley. J., 2009 Planetary Boundaries: Exploring the Safe Operating Space for Humanity. *Ecol. and Soc*. 14(2).

Richardson, Lewis Fry. 1960. *Statistics of Deadly Quarrels*. Pittsburgh, PA: Boxwood Press.

Ripple, William, Christopher Wolf, Thomas Newsome, Mauro Galetti, Mohammed Alamgir, Eileen Crist, Mahmoud Mahmoud, and William Laurence. 2017. World Scientists' Warning to Humanity: A Second Notice. http://scientistswarning.forestry.oregonstate.edu/sites/sw/files/Warning_article_with_supp_11-13-17.pdf

Russell, Stuart, Anthony Aguirre, Ariel Conn, and Max Tegmark. 2018. Why You Should Fear "Slaughterbots"—A Response. *IEEE Spectrum*. https://spectrum.ieee.org/automaton/robotics/artificial-intelligence/why-you-should-fear-slaughterbots-a-response.

Sandberg, Anders. 2014. Guesses. *Flickr*. https://www.flickr.com/photos/arenamontanus/14427926005/in/photolist-axC1R7-5LFwwU-7hQJqk-9bt9At-pY8ypH-nYWTJV-hS8HiV-kDqhCb-cxqxN1-cxqrZJ-pBqeDj-odfdF2-4DqCXj-f3rfff-mPsvky-6qqYwu-cSuQJu-c4jqZ5-6Jaj5R-9VYgo7-jzAnZC-gtTN7P-uZ3z1K-vHKd3U-qqsUqQ-7cUu4M/.

Sandberg, Anders. 2017 Existential Risk: How Threatened is Humanity? Presentation at Chalmers University of Technology.

Sandberg, Anders, and Nick Bostrom. 2008. Global Catastrophic Risks Survey, Technical Report #2008-1. https://www.fhi.ox.ac.uk/reports/2008-1.pdf.

Sandberg, Anders, and Nick Bostrom. 2011. Machine Intelligence Survey. FHI Technical Report. https://www.fhi.ox.ac.uk/wp-content/uploads/2011-1.pdf.

Sandberg, Anders, Eric Drexler, and Toby Ord. 2018. Dissolving the Fermi Paradox. arXiv. https://arxiv.org/pdf/1806.02404.pdf.

Satish, Usha, Mark J. Mendell, Krishnamurthy Shekhar, Toshifumi Hotchi, Douglas Sullivan, Siegfried Streufert, and William J. Fisk. 2012. Is CO_2 an Indoor Pollutant? Direct Effects of Low-to-Moderate CO2 Concentrations on Human Decision-Making Performance. *Environmental Health Perspectives*. 120(12): 1671-1677.

SAW. (2017). Slaughterbots. YouTube. https://www.youtube.com/watch?v=9CO6M2HsoIA&t=.

Scheelbeek, Pauline, Frances Bird, Hanna Tuomisto, Rosemary Green, Franscesca Harris, Edward Joy, Zaid Chalabi, Elizabeth Allen, Andy Haines, and Allan Dangour. 2018. Effect of Environmental Changes on Vegetable and Legume Yields and Nutritional Quality. *PNAS*. 115(26): 6804-6809.

Scheffler, Samuel. 2016. *Death and the Afterlife*. Oxford: Oxford University Press.

Scheffler, Samuel. 2018. *Why Worry About Future Generations?* Oxford: Oxford University Press.

Scranton, Roy. 2013. Learning How to Die in the Anthropocene. *New York Times*. https://opinionator.blogs.nytimes.com/2013/11/10/learning-how-to-die-in-the-anthropocene/#more-150341.

SD. 2015. Failing Phytoplankton, Failing Oxygen: Global Warming Disaster Could Suffocate Life on Planet Earth, ScienceDaily. https://www.sciencedaily.com/releases/2015/12/151201094120.htm.

Sekerci, H., Petrovskii, S., 2015 Mathematical Modeling of Plankton-Oxygen Dynamics Under the Climate Change, B. Math. Biol. 77(12): 2325-2353.

Senthilingam, Meera. 2017. Seven Reasons We're at More Risk than Ever of a Global Pandemic. CNN. http://www.cnn. com/2017/04/03/health/pandemic-risk-virus-bacteria/index.html.

Shimizu, Mika, and Allen Clark. 2015. Interconnected Risks, Cascading Disasters, and Disaster Management Policy: A Gap Analysis. *GRF Davos, Planet at Risk*. 3(2): 260-270.

Snopes. 2017. Did President Trump Reverse an Insecticide Ban After Receiving $1 Million from Dow Chemicals? *Snopes*. https://www.snopes.com/fact-check/trump-reverses-insecticide-ban-dow-chemicals/.

Snyder, R. 2016 A Proliferation Assessment of Third Generation Laser Uranium Enrichment Technology, Sci. Glob. Secur. 24(2): 68-91.

Sotala, Kaj, and Roman Yampolskiy. 2014. Responses to Catastrophic AGI Risk: A Survey. *Physica Scripta*. 90(1): 1-33.

Sotos, J. 2017 Biotechnology and the Lifetime of Technical Civilizations, arXiv.org. https://arxiv.org/abs/1709.01149.

Spaaij, Ramón. 2010. The Enigma of Lone Wolf Terrorism: An Assessment. *Studies in Conflict and Terrorism*. 33(9): 854-870.

Stern, N. 2006. Stern Review on the Economics of Climate Change. https://www.webcitation.org/5nCeyEYJr?url=http://www.hm-treasury.gov.uk/sternreview_index.htm.

Stout, Martha. 2005. *The Sociopath Next Door*. New York, NY: Broadway Books.

Taleb, Nassim. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York, NY: Random House.

Tegmark, Max. 2016. The Wisdom Race Is Heating Up. edge.org. https://www.edge.org/response-detail/26687.

Tegmark, Max. 2018. The Top Myths about Advanced AI. Future of Life Institute. https://futureoflife.org/background/aimyths/.

Tigchelaar, Michelle, David Barttisti, Rosamond Naylor, and Deepack Ray. 2018. Future Warming Increases Probability of Globally Synchronized Maize Production Shocks. *PNAS*. 115(26): 6644-6649.

Todd, Benjamin. 2017. Why Despite Global Progress, Humanity Is Probably Facing its Most Dangerous Time Ever. 80000 Hours. https://80000hours.org/articles/extinction-risk/.

Tomasik, Brian. 2017. Risks of Astronomical Future Suffering, Foundational Research Institute. https://foundational-research.org/risks-of-astronomical-future-suffering/.

Tonn, Bruce. 2009. Beliefs about Human Extinction. *Futures*. 41(10): 766-773.

Towers, Sherry, Andres Gomez-Lievano, Maryam Khan, Anuj Mubayi, and Carlos Castillo-Chavez. 2015. Contagion in Mass Killings and School Shootings. *PLOS*. 10(7).

UN. 2017. World Population Prospects. https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_Key-Findings.pdf.

Verdoux, Philippe. 2009. Transhumanism, Progress, and the Future. *Journal of Evolution and Technology*. 20(2): 49-69.

Verdoux, Philippe. 2011. Emerging Technologies and the Future of Philosophy. *Metaphilosophy*. 42(5): 682-707.

Walker, Mark. 2002. Prolegomena to Any Future Philosophy. *Journal of Evolution and Technology*. 10.

Wells, Willard. 2009. *Apocalypse When?: Calculating How Long the Human Race Will Survive*. New York, NY: Springer Praxis Books.

Wiblin, Robert. 2017. Why the Long-Term Future of Humanity Matters More Than Anything Else, and What We Should Do About It. 80,000 Hours Podcast. https://80000hours.org/podcast/episodes/why-the-long-run-future-matters-more-than-anything-else-and-what-we-should-do-about-it/.

Willett, K., Sherwood, S., 2012 Exceedance of Heat Index Thresholds for 15 Regions Under a Warming Climate Using the Wet-Bulb Globe Temperature, Int. J. of Clim. 32(2): 161-177.

Williams, Christopher. 1997. *Terminus Brain. The Environmental Threats to Human Intelligence*. London, UK: Cassel, London.

Wilson, EO. 2006. *Nature Revealed: Selected Writings, 1949-2006*. Baltimore, MD: Johns Hopkins University Press.

Woolley, Anita, Christopher Chabris, Alex Pentland, Nada Hashmi, and Thomas Malone. 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science*. 330(6004): 686-688.

Worm, B., Barbier, E., Beaumont, N., Duffy, J.E., Folk, C., Halpern, B., Jackson, J., Lotze, H.K., Micheli, F., Palombi, S., Sala, E., Selkoe, K., Stachowicz, J., Watson, R. 2006 Impacts on Biodiversity Loss on Ocean Ecosystem Services, Science 314: 787-790.

WWF. 2014. Living Planet Report. http://awsassets.panda.org/downloads/lpr_living_planet_report_2014.pdf.

Yampolskiy, Roman. 2016. *Artificial Superintelligence: A Futuristic Approach*. New York, NY: Taylor and Francis Group.

Yudkowsky, Eliezer. 2008a. Cognitive Biases Potentially Affecting Judgement of Global Risks. In Nick Bostrom and Milan Ćirković (eds.), *Global Catastrophic Risks*. Oxford: Oxford University Press.

Yudkowsky, Eliezer. 2008b. Artificial Intelligence as a Positive and Negative Factor in Global Risk. In Nick Bostrom and Milan Ćirković (eds.), *Global Catastrophic Risks*. Oxford: Oxford University Press.

Zimmer, Shanta, and Donald Burke. 2009. Historical Perspective—Emergence of Influenza A (H1N1) Viruses. *New England Journal of Medicine*. 361: 279-285.