Facing Disaster: The Great Challenges Framework

Phil Torres

Keywords: existential risk, global catastrophic risk, environmental degradation, emerging technologies, artificial intelligence

This paper offers a comprehensive overview of our evolving existential predicament. Section 1 provides some background to the topic. Section 2 examines the three "Great Challenges" facing humanity, namely, environmental ruination, the distribution of unprecedented destructive capabilities across society, and value-misaligned machine superintelligence. Section 3 considers some objections and complications to the Great Challenges framework. The final section observes that if we consider the emergence of recorded history about 4,000 years ago to be the beginning of civilization, then we may have progressed through a mere 0.0002 percent of our entire potential history on Earth; mapping this onto the annual calendar, we are no more than about 63 seconds into the first day of the year. This makes the study and mitigation of "existential risks" one of the most important tasks for humanity this century.

1. Introduction

After some 188,000 years of living in nomadic, hunter-gatherer communities no larger than about Dunbar's number (~150), our ancestors decided to conduct an unscientific experiment: *civilization*. So far, this experiment has had mixed results. As Bertrand Russell writes in his reflection on humanity's future, "Prologue or Epilogue?" (1954), "progress has not been steady, but has been a matter of fits and starts." For example, consider that the Neolithic revolution resulted in an appreciable loss of average human height due to poorer diets; humanity only regained our "natural" average height in the middle of the twenty-first century (Cohen 1991). Similarly, life expectancy at birth fell from 33 to as low as 20 years. The creation of social hierarchies, which replaced the egalitarianism of previous communities, also enabled new forms of oppression. And many pathological conditions faced by humans today, from diabetes and obesity to chronic anxiety and insomnia, are classifiable as "diseases of civilization," since they are the result of occupying social and physical environments for which natural selection did not "design" us (see Jacobs 2003).[1] Furthermore, if one were to add up all the human suffering in the world today and compare it to the total amount of suffering experienced at any given timeslice by our *Homo* ancestors, the mere fact that our population has ballooned from ~4 million circa the Neolithic revolution to ~7.6 billion entails that the contemporary world contains far more human misery than ever before

Email address: philosophytorres@gmail.com (Phil Torres)

[1] Diseases of civilization should be distinguished from "diseases of longevity," which are the result of living longer.

February 14, 2018

in history. This point is especially significant if one adheres to a "suffering-focused" ethics (Gloor and Mannino 2018).

But there have also been tremendous strides toward human betterment. We no longer worry about dying from medical conditions like appendicitis. The opportunities for self-actualization are truly unprecedented. And our understanding of the nature and workings of reality—the arcana of this strange cosmic abode—has never been more complete. Even more, studies suggest a transhistorical decline in violence from the Paleolithic to the present. Today, humanity finds itself in the midst of the Long Peace, a period during which no two world superpowers have gone to war. On this model, Steven Pinker (2011) coins the term "New Peace" to describe the period from the end of the Cold War to now, during which "organized conflicts of all kinds—civil wars, genocides, repression by autocratic governments, and terrorist attacks—have declined throughout the world." There have also been Rights Revolutions that have brought about attitudinal changes about the acceptability of violence toward "ethnic minorities, women, children, homosexuals, and animals." Pinker argues that moral progress in the past few centuries has been driven by rising IQs around the world, i.e., the "Flynn effect," which has increased our capacity for abstract reasoning, in turn leading to an expansion of our "circles of moral concern"—a phenomenon that Pinker calls the "moral Flynn effect" (Pinker 2011). From this perspective, the present is far brighter than the past, and the future may be brighter still, if these trends continue. Other scholars associated with the "New Optimists" label are Johan Norberg, Hans Rosling, Max Roser, Nicholas Kristof, and Matt Ridley.

Yet I would argue that the way this form of Enlightenment progressionism is often presented is misleading. While some scholars have questioned the statistical analyses underlying Pinker's thesis (see, e.g., Price 2017; Taleb 2016), the more egregious error is its unbalanced focus on what is going right and often unfair dismissal of the formidability and anomalousness of the global-scale problems facing humanity today and in the coming century. Bill Clinton is fond of saying, "follow the trendlines, not the headlines," but it also matters *which* trendlines one follows (Torres 2016a). A more complete picture of the human situation at the mid-morning of the twenty-first century not only recognizes the immense moral, scientific, material, and so on, advances, but that our world contains far more *risk potential* than ever before in the 2.5-million-year history of our taxonomic genus. "Risk potential" is a measure of the possible harm that humanity could encounter. There are two related trends to consider here.

First, the absolute *number* of global-scale catastrophe scenarios has significantly risen since the middle of last century (Torres 2016b, 2017a). Prior to this moment, the only risks to our survival stemmed from natural phenomena like asteroids, comets, supervolcanoes, gamma-ray bursts, and pandemics. Today the working list of "existing" and "emerging" risks includes not just these, but anthropogenic phenomena like climate change, biodiversity loss, species extinctions, nuclear conflict, designer pathogens, atomically-precise manufacturing, autonomous nanobots, geoengineering, physics disasters, and machine superintelligence—to name a few. And we should not forget the possibility that future technologies currently hidden beneath the horizon of our collective imagination could introduce entirely new risks to our survival. Elsewhere I have argued that "unknown unknowns"—a term that could

subsume novel inventions, unintended consequences from purposive action, or natural phenomena of which we are currently ignorant—could constitute the greatest long-term threat (Torres 2017a). If the overall threat grows in proportion to the exponential development of technology, then we might even consider talking about an "existential risk singularity" (Verdoux 2009).

Second, this growing list has led many experts to assign an unsettlingly high subjective *probability* to a global catastrophe occurring. On the one hand, the probability of annihilation per century from our "cosmic risk background" is almost certainly less than 1 percent; Toby Ord (2015) argues that it may be much less than 1 percent. By contrast, Sir Nicholas Stern assumes a 0.01 percent chance of extinction each year in his influential Stern Report (2006), which adds up to a 9.5 percent chance per century. But Stern chose this number as a modeling assumption for the purpose of discussing discount rates. When one seriously considers the growing swarm of doomsday scenarios, though, it appears that the probability could be far higher. Consider the following:

1. John Leslie estimates that the probability of annihilation in the next 500 years is 30 percent (Leslie 1996).[2]
2. The director of FHI, Nick Bostrom, writes that his "subjective opinion is that setting this probability lower than 25 percent would be misguided, and the best estimate may be considerably higher" (Bostrom 2002).
3. Lord Martin Rees puts the likelihood of civilizational collapse before 2100 at 50 percent (Rees 2003).
4. Richard Posner judges the near-term chance of extinction to be "significant," adding that "human extinction is becoming a feasible scientific project" (Posner 2004).
5. An informal survey of experts conducted by the Future of Humanity Institute (FHI) yielded a probability of extinction this century at 19 percent (Sandberg and Bostrom 2008).
6. 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).
7. Finally, 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).[3]

If one takes these estimates seriously—and we will provide reasons below for doing so—they imply that the average American is literally *thousands* of times more likely to encounter

[2]Note that this is based partly on the Doomsday Argument mentioned at the end of section 3.

[3]Others have made similar conjectures; for example, Frank Fenner speculated in 2010 that "humans will probably be extinct within 100 years" (Edwards 2010); Guy McPherson claims that humanity will die out by 2026 (Smallman 2016); and Neil Dawe says that "he wouldn't be surprised if the generation after him witnessed the extinction of humanity" (Jamail 2013). On the other hand, Richard Gott uses statistical analyses to estimate that humanity will go extinct between 5,100 and 7.8 million years from today.

a civilization-ending disaster than, say, die in an air and space transport accident. For example, the FHI survey suggests that the average American is at least 1,500 times more likely to perish in a human extinction catastrophe 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 good chance of living to see an existential catastrophe of some sort (Torres 2016c, 2017a; Wells 2009).[4]

These estimates are also consistent with warnings made by a number of leading intellectuals around the world. For example, Stephen Hawking writes in a *Guardian* op-ed "that we are at the most dangerous moment in the development of humanity" (Hawking 2016), later suggesting that our species has about 100 years to leave planet Earth "or die" (Fecht 2017). Similarly, Noam Chomsky has argued on numerous occasions that the risk of human annihilation is "unprecedented in the history of *Homo sapiens*" (Lombroso 2016). And Ingmar Persson and Julian Savulescu (2012) argue that "our present situation is so desperate" that humanity should consider the use of radical moral bioenhancements that augment our moral dispositions of altruism and the sense of justice. More recently, within the arena of geopolitics, the UN Secretary-General Antonio Guterres ended 2017 with an unprecedented warning: "On New Year's Day 2018, I am not issuing an appeal, I am issuing an alert—a red alert for our world" (Samuels 2017). This ominous declaration followed a short paper published a few months earlier and signed by over 15,000 scientists, titled "World Scientists' Warning to Humanity: A Second Notice." Focusing on environmental ruination in particular, it states that "humanity has failed to make sufficient progress in generally solving . . . foreseen environmental challenges, and alarmingly, most of them are getting far worse." It concludes 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).

Our profound impact on the world has also led scientists to propose a new geological epoch, the "Anthropocene," whose signatures include climatic and biospheric disruptions, as well as the global distribution of artificial radionuclides from thermonuclear testing. To quote Jennifer Jacquet (2017), "not since cyanobacteria has a single taxonomic group been so in charge" of crucial planetary systems. Finally, consider the most chilling lesson of the Fermi paradox, namely, that *species virtually never make it past our current level of technological development.* The reason is that if they did, we would likely see (some trace of) astrobiological evidence of them; yet the cosmos appears lifeless and barren in all directions.

[4]Or, as Wells (2009) puts it, "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."

We appear to be alone, and one prominent explanation is the "doomsday hypothesis," i.e., that technologically advanced civilizations tend to self-destruct just as they develop the capacity for space colonization. In fact, E.O. Wilson (2006) and Ernst Mayr (1995) have both (independently) suggested, from an evolutionary biology perspective, that intelligence may be something akin to a "lethal mutation," thus explaining the deafening quietude of the "Great Silence" (Brin 1983).[5]

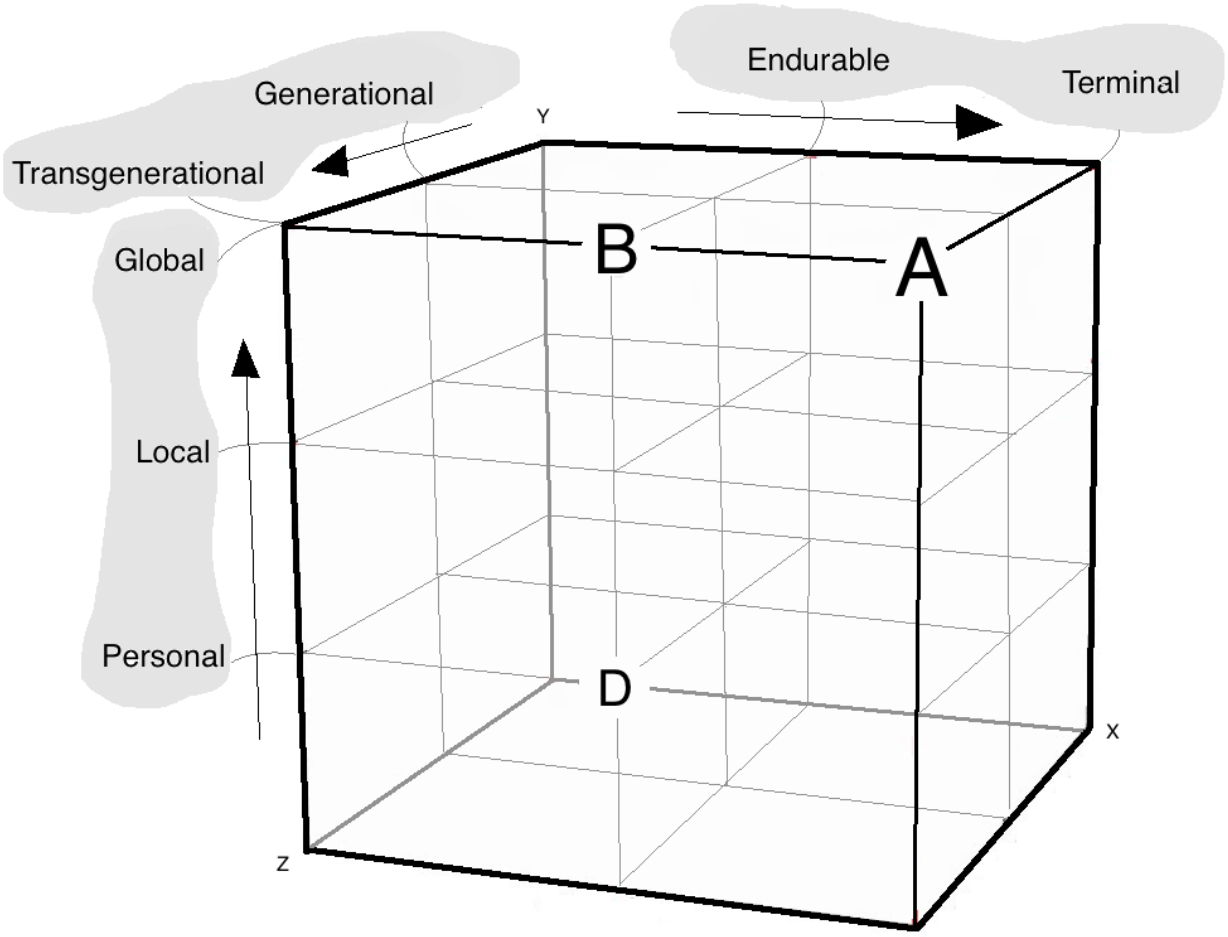

Figure 1: The x-axis represents severity, the y-axis represents spatial scope, and the z-axis represents temporal scope. Risk outcomes generally become worse as one follows the arrows (Torres 2017a).

With this backdrop in place, the present paper offers a comprehensive survey of our evolving existential predicament. In particular, it adopts what I will call the "Great Challenges framework," which is designed to provide a roadmap to help civilization navigate the

[5]The specific term "lethal mutation" comes from Chomsky (2010).

obstacle course of risks before us. For a phenomenon to qualify as a "Great Challenge," it must satisfy three criteria: (i) significance; it must have existential risk implications, (ii) urgency; it must require our immediate attention if we are to have a good chance of neutralizing it, and (iii) inevitability; it must be the case that it is very likely to be encountered, given the current trajectory of our civilization's development. The concept is intended mostly as a useful organizing principle rather than as a strict ontological category like "hydrogen" or "tensor." It is "useful" because it thrusts into the foreground a number of problems that, insofar as one accepts that existential risks are uniquely morally significant, ought to be prioritized over all (or at least most) others (see Bostrom 2013). Here we will define an existential risk as any event that either trips our species into the eternal grave of extinction or permanently prevents us from exploiting a sizable portion of our cosmic endowment of negentropy to realize astronomical amounts of value. With respect to Figure 1, all node-A events are existential catastrophes, whereas any node-B event that satisfies the second disjunct above constitutes an existential catastrophe. 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), as well as the present author (see Appendix 1), provide a rough outline of risk outcomes organized by their causes (from criticisms of this approach, see Häggström 2016).

In the following section, we will explore three Great Challenges, namely, environmental degradation, the democratization of science and technology (i.e., the distribution of unprecedented offensive capabilities across society), and superintelligence. The penultimate section will then explore a number of objections and additional issues relating to this framework.

2. Three Great Challenges

Let's examine the Great Challenges in turn:

2.1 Environmental Degradation. 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 (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 follow (Willett and Sherwood 2012). 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, sci-

entists 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.

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. These studies 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 continues to question climate science, rose to power—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, an analysis of the Malthusian problem of the population growing faster than its food supply found

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 tolerance of animals . . . may occur by ~2200-2400, independent of the predicted slowdown in population growth by 2100 (Haqq-Misra at al. 2017).

Anthropogenic carbon dioxide 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 was 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 over 500 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 2017a).

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). Another study found that current trends, when extrapolated into the future, imply that 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 they put it.

But the loss 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 within the tropics—that is, between the Tropic of Cancer and the Tropic of Capricorn—fell by an incredible 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."[6] Similarly, the 2016 Living Planet Report states that the global abundance of wild vertebrates declined by a staggering 58 percent between 1970 and 2012, and we could witness a decline of two-thirds by 2020 (WWF 2014). 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 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* (quoted in Torres 2017a). This has implications for agricultural production, which is especially unsettling 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 a population 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).

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" that 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. It 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

[6]Notice that people often don't consider how human activity might be affecting plant life, a phenomenon that scientists call "plant blindness."

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 will almost certainly be our greatest legacy on the planet. This 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 (Torres 2017a). As the authors write, 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.

These phenomena pose direct threats to the prosperity and perpetuation of our species. Many civilizations throughout history, such as the Maya and Rapa Nui civilizations, have collapsed due to environmental degradation caused by deforestation, pollution, overfishing, and so on (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—to falling victim to what Daniel O'Leary (2006) calls a "progress trap." In fact, 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.

This is far from an exhaustive survey of the environmental problems facing humanity today. Other issues that could be elaborated on include the build-up of plastic trash in the ocean's gyres, the greening of the Antarctic, the "zombie pathogens" emerging from thawing permafrost, the creeping forward of Overshoot Day, the protracted degradation period of plastics and glass, and the connection between CO2 levels and cognitive performance (see

section 3). Suffice it to say that environmental degradation is a significant, urgent, and unavoidable challenge that humanity will have to confront and solve if we wish to survive on Planet A, our pale blue dot.

2.2 The Democratization of Science and Technology. Prior to 1945, no actor had the capacity to unilaterally destroy the world. With the commencement of the Atomic Age, two state actors gained this unique ability. Today, techno-developmental trends suggest that the number of state and nonstate actors who could cripple civilization is increasing. There are three features, in particular, of emerging technologies that one must understand to appreciate the unprecedented dangers posed by this rapidly changing situation. By "emerging technologies," we primarily mean those instruments, techniques, and artifacts associated with biotechnology, synthetic biology, nanotechnology (especially atomically-precise manufacturing), and artificial intelligence.

1. *Use-Flexibility.* Emerging technologies are *dually usable*, meaning that the very same artifact can be used for both beneficial and harmful ends. A centrifuge that can enrich uranium for nuclear power plants can also be used to enrich it for nuclear weapons; a laboratory that could find a cure for Ebola could also be used to weaponize the virus.
2. *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, the Carlson curve, Dennard scaling, Keck's law, Kryder's law, and other trends broadly subsumed under the Kurzweilian "Law of Accelerating Returns" (Kurzweil 2005).
3. *Democratization.* Emerging technologies are increasingly *accessible* to small groups, lone wolves, and "lone wolf packs" (Pantucci 2011). Examples include CRISPR/Cas-9, digital-to-biological converters, base editing, USB-powered DNA sequencers, SILEX (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 weaponry; autonomous nanobots that could target specific people, races, or species; lethal autonomous drones—e.g., "slaughterbots" (Russell et al. 2018)[7]—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 metamaterial invisibility cloaks, self-guided bullets, cognitive enhancements, exoskeletons, robot soldiers, direct-energy weapons (DEWs) like laser and particle beam weapons, and mind-control/mind-reading technologies to further complicate the situation.

It is important to recognize that the technologies listed above could bring about truly marvelous improvements in the human condition by eliminating disease, reversing aging, and enhancing morality (Kurzweil 2005; Diamandis 2014; Persson and Savulescu 2012). Yet the property of (i) entails that these very same inventions will simultaneously engender profound

[7]See also SAW 2017.

risks to our collective well-being and survival, perhaps resulting in a large and irreversible loss of expected value (Cotton-Barratt and Ord 2015). Given a conception of the 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, scholars have recently outlined a quadripartite typology of human agents who are prime candidates for intentionally bringing about an existential catastrophe, if only the means were available to them. These are: (1) *Apocalyptic terrorists*, i.e., religious extremists who believe that the world must be destroyed to be saved. (2) *Misguided moral actors*, e.g., radical negative utilitarians who advocate annihilation to eliminate suffering. (3) *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). To convey a sense of how dangerous some of these individuals could be if synthesizing pathogens becomes as easy as obtaining a bump stock or Kalashnikov, consider a few quotes from individuals in categories (3) and (4) (the first two categories require more context to understand).

In the *Earth First! Journal*, an anonymous author writes the following:

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. Remember, Equal Rights for All Other Species (Dye 1993).

This pro-omnicide view is common among individuals within the most radical fringe of the 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 voluntarily or involuntarily through some global-scale catastrophe like an engineered pandemic (quoted in Torres 2017b, 2018a). 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," adding, "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" (quoted in Torres 2017b, 2018a). And finally, the 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).

Many individuals of this sort 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). While not all sociopaths are sadistic or violent, they do comprise a disproportionate percentage of the prison population—about 20 percent (Torres 2017a). 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 people will pose an increasingly significant threat—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 in the past several decades (Miller 2015; Jenkins et al. 2016).

With this in mind, let's quantify the hypothetical level of threat that terror agents could pose, I propose the *AW formula*, where "AW" stands for the necessary and sufficient conditions, "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. 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 doom 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 eveaen 1—percent, given the trend of (iii), as well as 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 catastrophe could be more or less inescapable.

The situation is even worse than this, though, because of agential error. This 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, whereas relatively few of the latter instantiate the former. Here we can use the *EA formula* to estimate the danger posed by fallible humans 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. First, although one might—naively—surmise that this probability would be extremely low, the history of mistakes with real-world consequences, even in highly regulated government or university laboratories, is extensive and sobering. For example, the 2009 swine flu outbreak, which may have killed some 203,000 people around the world, was likely the result of a virus released from a laboratory in the late 1970s; a government report specifies more than 1,1000

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 2016b, 2017a). Even more, consider that the community of professional geneticists is rapidly growing: roughly 1.5 million scientists published scientific articles indexed under "genetic techniques" between 2008 and 2015 (Sotos 2017). Even more, the biohacker movement is rapidly growing as well, as mail-order CRISPR kits and other DIY equipment become more and more affordable to interested hobbyists. Thus, the overall probability of *someone somewhere* making a mistake with potentially catastrophic consequences may not be small.

This has ominous implications. For example, imagine a future in which only 500 of 10 billion people—just 0.000005 percent of the population—have access to world-destroying technologies. We can stipulate that these are uniformly well-intentioned individuals. If each person has a mere 0.01 chance of accidentally initiating a doomsday disaster per decade, the probability of self-annihilation would be 99.3 percent. One gets even more certain doom 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. Thus, once again it appears that the trend (iii) plus the trend of (ii) and the property of (i) may pose an exceptionally high risk to our collective survival. Also note that whereas the discussion above has focused on biotechnology and synthetic biology, the advent of atomically-precise nanotechnology, autonomous drones, and so on, could place *even more* dual-use power in the hands of *even more* individuals, thus multiplying the total number of token agents with the unilateral capacity to wreak unprecedented havoc on civilization.

John Sotos (2017) proposes a similar model that yields comparably pessimistic results. Focusing on biotechnology, he calculates that a 1 in 100 chance of only a few hundred 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 using 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 2). 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%." Sotos concludes that his calculations supply "the quantitative 24 orders-of-magnitude winnowing required of a Great Filter," given that the visible universe contains "$\approx 10^{24}$ 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). Incidentally, 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).

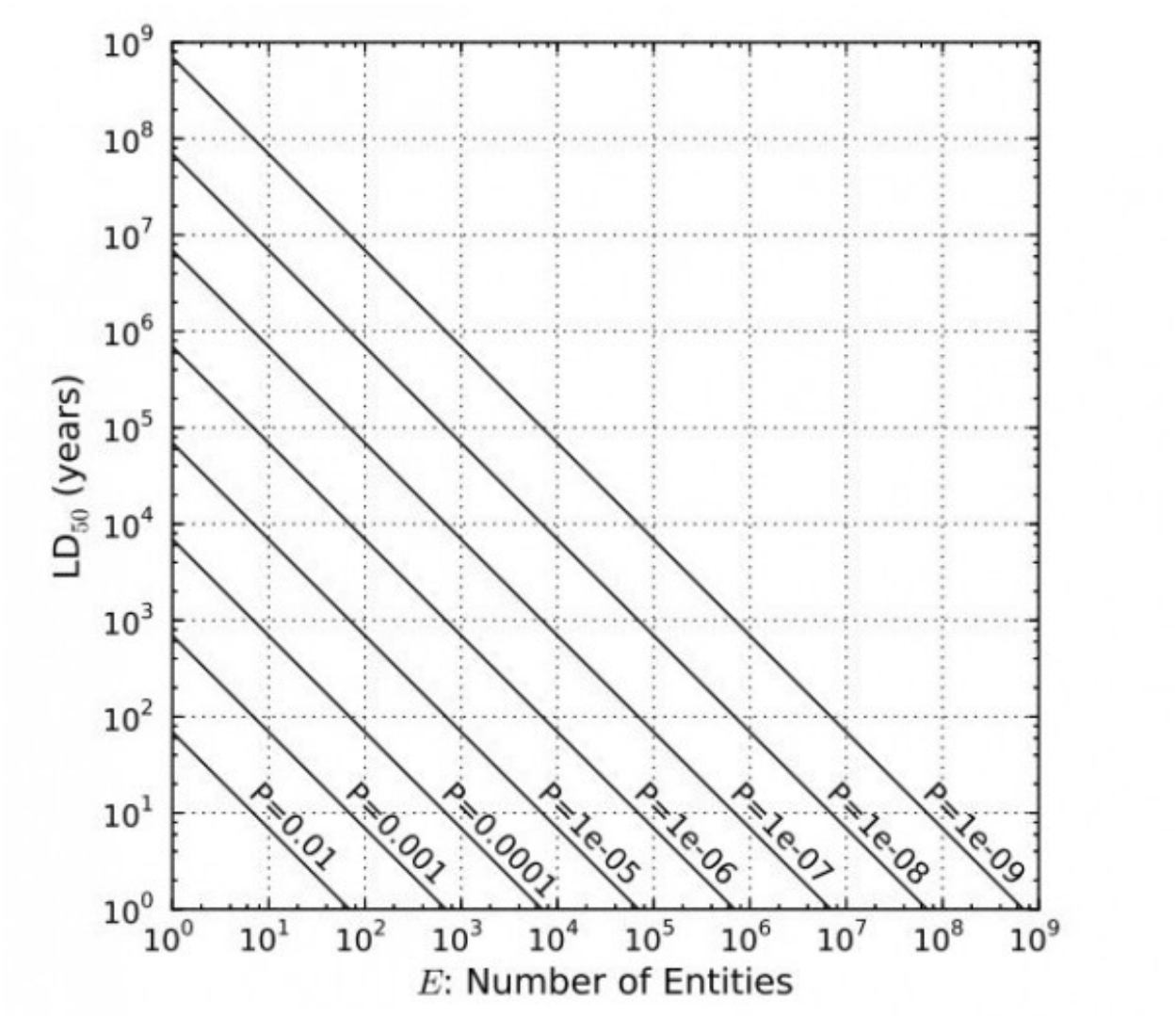

Figure 2: From Sotos (2017).

Finally, it may be worth noting some estimates of the likelihood of one risk in particular, namely, that associated with nuclear weapons. Anders Sandberg (2014) identifies this as the greatest existing (rather than emerging) threat to humanity. For example, Robert Galluccii stated in 2005 that "it is more likely than not that al-Qaeda or one of its affiliates will detonate a nuclear weapon in a US city within the next five to ten years" (Gallucci 2005), and Graham Allison (2004a, 2004b) conjectured that a nuclear attack in America between 2004 and 2014 was "more likely than not." An equally pessimistic figure comes from a 2008 report issued by the US Commission on the Prevention of Weapons of Mass Destruction Proliferation and Terrorism, which concludes that "it is more likely than not that a weapon of mass destruction will be used in a terrorist attack somewhere in the world by the end

of 2013" (Graham et al. 2008). More generally, a survey of 85 national security experts in 2005 found that "60 percent of the respondents assessed the odds of a nuclear attack within 10 years at between 10 and 50 percent, with an average of 29.2 percent" (see Torres 2016b). Almost 80 percent of those polled said the attack would come from terrorists. Finally, Martin Hellman puts the probability of a nuclear bomb being detonated at 1 percent every year from the present, meaning that "in 10 years the likelihood is almost 10 percent, and in 50 years 40 percent if there is no substantial change" (Farber 2010). In other words, *it is almost certain that someone will detonate a nuclear weapon by the end of this century*, a conclusion that is consisted with Lawrence Krauss's assertion: "I think the use of nuclear weapons is [inevitable]. I think having a nuclear weapon used by accident or on purpose against a civilian population is unavoidable as long as we possess them" (Torres 2017c).

It appears that the cliche, "it's a matter of when rather than if," applies in this context. The number of potential perpetrators is growing too large for humanity to escape a disaster—at least from the evidential vantage point of the present moment.

2.3 Value-Misaligned Machine Superintelligence. Many existential risk scholars, including the present author, believe that superintelligence 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." 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, with 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. Here we can dismiss criticisms of "Friendly AI" safety engineering projects 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, but rather that its goals could be misaligned with ours *and* it could be competent enough to achieve them. In more detail, the dangers posed by superintelligence stem can be dissected into the following six theses:

1. *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 genuine superintelligence that only values playing tic-tac-toe, writing and rewriting Shakespeare's plays, or worshiping some AI god that it "believes" in.

2. *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 maximizing the total number of paperclips in the universe, calculating as many digits of pi as possible, proving Goldbach's conjecture, and so on.
3. *Complexity of value thesis.* Our values—i.e., the motivating values we would load into an artificial intelligence—have a high Kolmogorov complexity. That is to say, they cannot be reduced to a simple codifiable list of prescriptions and proscriptions.
4. *Fragility of value thesis.* It appears that successfully loading most of our values into an artificial intelligence will 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).
5. *Relative speed thesis.* The electrical potentials within current 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. 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 us. And finally,
6. *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 the dung beetle.

There are 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 incredibly high-stakes problems: according to a recent survey of AI experts, *nearly everyone surveyed* said that a human-level AI will join humanity by 2100 (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—meaning that there is a high probability, insofar as expert opinion is credible, for expecting a superintelligence to join humanity by 2100

(see Armstrong and Sotala 2012). Another survey found that 75 percent of respondents—all fellows of the Association for the Advancement of Artificial Intelligence (AAAI)—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 "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."[8]

So, it appears that by the end of this century—if not within a few decades, if not within a few years—we will 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. The first could ultimately be insoluble. Just consider the vast range of values that Christians, Catholics, Muslims, Hindus, Buddhists, agnostics, atheists, anti-theists, Republicans, Democrats, libertarians, socialists, communists, fascists, nationalists, populists, monarchists, anarchists, white supremacists, civil rights activists, men's rights advocates, feminists, bioconservatives, transhumanists, pronatalists, antinatalists, *and so on*, accept. 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 and 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, a leading figure in AI development, 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 is will continue to accelerate in the coming years and decades—that is, in the absence of a defeater like an existential catastrophe—this risk

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

appears to be unavoidable. 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 definition.[9]

* * *

Finally, we can call the first challenge of this section a *context risk* because, while environmental degradation poses many direct threats to human well-being on a global scale, it also *frames* our more general existential predicament on Earth (Torres 2017a). As such, it has the capacity to *modulate* the risks associated with interstate conflicts, civil wars, terrorist attacks, and other forms of violence, and for this reason I believe that context risks may be the most urgent of all existential risks here adumbrated.[10] In contrast, the distribution of unprecedented offensive capabilities across society is a *state risk*, or a risk associated with being in some particular state or configuration—in this case, the configuration of many actors with the unilateral capacity to inflict civilization-ending harm. Consistent with the calculations of section 2, the longer one is exposed to a state risk, the higher the likelihood of disaster. Lastly, superintelligence seems to constitute a *step risk*, or a risk 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, 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 prioritize one over the others.

3. Objections and Complications

Consider some additional considerations that are relevant to understanding our current existential predicament.

(i) One might object to some of the doomsday scenarios outlined above, e.g., global pandemics and nuclear winters, by citing the *last few people problem*: one can readily devise hypothetical narratives in which a large number of humans perish in a catastrophe, but it becomes much more difficult to imagine how the "last few people" might follow their conspecifics to the grave (Tonn and MacGreagor 2009; Torres 2017a). In the case of a pandemic, consider the uncontacted tribes in the Brazilian Amazon, the 80 to 150 people stationed in Villa Las Estrellas, and the many personnel safely contained in the aquatic refuges of submarines (see Turchin and Green 2017). How does a pathogen wipe out such individuals? There are three responses here: first, some omnicidal groups, such as GLF (mentioned in section 2.2), have explicitly discussed the possibility of releasing multiple pathogens sequentially to ensure that political leaders, etc. will die once they emerge from their bunkers (see Torres 2017b, 2018a). And second, even if some people were to survive,

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

[10]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."

their geographical distribution could prevent them from generating sufficient genetic diversity of offspring to ensure the survival of humanity. There would need to be about 1,000 breeding individuals in close proximity to perpetuate the species. And third, one doesn't need to cause human extinction to cause an existential catastrophe: it could be the case that wiping out all but 1 percent of the current human population—meaning that 75 million people survive—is enough to permanently stifle civilization, set us on a course of "recurrent collapse," or whatever (Bostrom 2013).

Furthermore, some risk scenarios could easily overcome the last few people problem. For example, a runaway greenhouse effect, self-replicating nanobots, superintelligence, and many physics disaster scenarios appear to pose all-or-nothing risks. If any of these risks were realized, the entire human population would perish.

(ii) One might also argue that the distribution of offensive capabilities across society could be overcome by some form of state surveillance. This is, in fact, what Ingmar Persson and Julian Savulescu advocate to reduce the dangers posed by terrorist groups and lone wolf attacks (Persson and Savulescu 2012). Such surveillance would need to be intrusive and asymmetrical, since a "transparent society," as envisaged by David Brin (1998), would give perpetrators too much information about law enforcement operations, thereby undercutting their efficacy. The obvious problem is the possibility of rogue officials or hackers misusing the information gathered for nefarious purposes. Or worse, such systems could be exploited by autocrats to oppress a population—a potential existential catastrophe in its own right, if the regime becomes sufficiently entrenched and global. Humans are far too venal, far too corruptible, to handle such power responsibly, and this makes intrusive, asymmetrical state surveillance extremely risky (Torres 2018b).

But there is an even more pressing problem associated with the dual usability, growing power, and increasing accessibility of emerging technologies. For states to provide security, they need a Weberian "monopoly" of force, power, and violence. Without this monopoly, efforts to impose law and order will be ineffectual, resulting in a state of Hobbesian anarchy. The crucial idea here is that the democratization of science and technology is empowering nonstate actors *more* than state actors (see Figure 3). When individuals have the capacity to unilaterally bring about global-scale catastrophes, the social contract cannot survive because states can no longer guarantee security for their citizens. Consequently, as Benjamin Wittes and Gabriella Blum (2015) suggest, we will need new forms of governance to survive. But what exactly might a new governing system look like? What form would it take? I have elsewhere argued that we may need to begin designing superintelligent algorithms for the specific task of governing. If, as I have claimed above, intelligence yields power, then the superior intelligence of a supersmart algorithm could restore the "monopoly" needed for states to effectively provide their *raison d'être*, namely, security. Yet this proposal encounters two serious problems: first, as discussed above, the control problem is perhaps the most formidable problem that our species has ever had to solve, and second, we would need to create this superintelligence *before* the point in Figure 3 where the two trendlines converge. Given the exponential rate of emerging tech development, this point could be much sooner than when superintelligence would otherwise be developed, meaning that we might have even less time to solve the control problem (Torres 2018b).

February 14, 2018

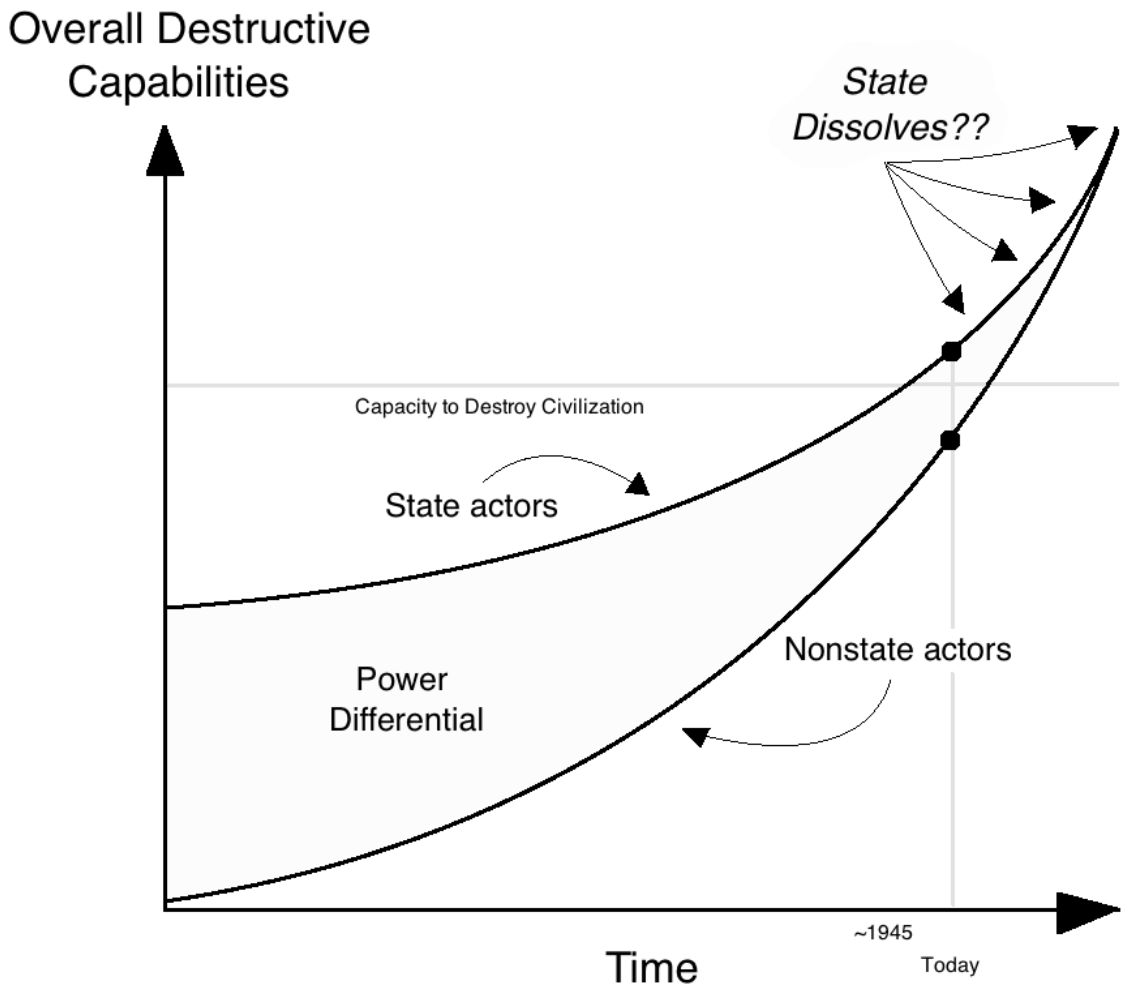

Figure 3: Schematic diagram. The democratization of science will undercut the social contract by reducing the power differential between state and nonstate actors (Torres 2018b).

Yet another problem concerns the increasing inability for political leaders to govern competently given the rapidly growing complexity of the world. According to Yaneer Bar-Yam, human civilization has complexified over time in part due to specialization, resulting in a gradual transformation from hierarchically-controlled systems to "networked" systems marked by lateral interactions between individuals and subgroups. The crucial idea is that in networked systems, no single individual or set of individuals can control, regulate, direct, or coordinate collective behaviors. As Bar-Yam (2002) puts it,

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.

It follows that civilization is reaching a point at which it cannot be effectively governed by political leaders and governmental agencies. In fact, this parallels a concern that is directly relevant to the field of existential risk studies: more than ever, we need generalists—highly capable generalists are called "polymaths"; less capable ones are called "dilettantes"—who can draw from a wide range of increasingly fragmented scientific and humanistic disciplines and, using the knowledge gathered, assemble a comprehensive, coherent "big-picture" map of where civilization is today and where it appears to be going. Yet this task is ever-more arduous given that, put somewhat crudely, information is the "fastest growing entity" in the universe (Kelly 2008). As another scholar 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 resulting "ignorance explosion" has yielded an epistemic predicament in which, sententiously put, "*everyone today knows almost nothing about most things*" (Verdoux 2011). This poses a dire conundrum for existential risk studies—how effective can such scholarship be given the growing differential between what even the smartest individuals know and what the collective as a whole "knows"?—as well as governments. In sum, ignorance leads to bad political leadership and it is the death knell of democracy, which depends upon an informed electorate. Yet in the absence of cognitive enhancement technologies that can significantly augment individual minds, unprecedented levels of individual ignorance are inevitable.

(iii) One might speculate that space colonization could significantly reduce the overall probability of existential risk. In fact, expanding into space is one of the primary reasons that, so far as I can tell, many futurological scholars accept what we can call the "bottleneck hypothesis."

Bottleneck hypothesis: Humanity finds itself in a unique moment of heightened hazards during which we are, or may be, unusually vulnerable to a global disaster; but if we play our cards right, the future could be brighter than ever before (Torres 2017a).[11]

The idea is that spreading into the three dimensions of space is the most cost-effective way of "playing our cards right." A list of notables who have held this view would include Elon Musk, Stephen Hawking, Derek Parfit, Nick Bostrom, Carl Sagan, Richard Gott, Jason Matheny, and Lord Martin Rees. But a closer look at what colonization would likely entail suggests a nontrivial, or perhaps high, probability that it would ultimately instantiate a "suffering risk" or "s-risk," i.e., a condition marked by "astronomical amounts" of suffering (Althaus and Gloor 2018; Tomasik 2017). To summarize the basic argument—at the loss of important argumentative detail, discussed elsewhere—expanding into space will generate a wide variety of distinct species, many of which will have their own cultural, political, reli-

[11] Toby Ord refers to this as the "risk window" idea, attributing it to Carl Sagan and John Leslie (among others) and describing it as follows: "We are in a very unusual historical period of high existential risk. This only started recently (with nuclear weapons) and will finish within a few centuries. It is thus a unique challenge in the history of our species."

gious, etc. traditions. Some of these species will be inclined to attack others for imperialistic reasons—broadly speaking, Hobbes's notion of "gain"—while everyone else will have strong game theoretic incentives to preemptively attack their neighbors for self-preservational reasons (the "Hobbesian trap"). This situation could be avoided if there were to exist a "cosmic Leviathan" or sorts, but this proposal appears unworkable given the hard constraints imposed on communication by the speed limit of light: a supreme governing system cannot be effective if it cannot coordinate its actions in a timely manner, and the limitations on physical travel and information exchange will make such coordination *untimely* across the vast expanses of space. Another possibility is that future civilizations implement policies of deterrence that establish a "balance of terror," which is (along with "pure dumb luck") what kept the Cold War *cold*. But this too appears implausible given the lightspeed action of advanced weaponry that technologically mature civilizations will almost certainly have at their disposal, such as heliobeams, direct-energy weapons, gravity wave guns, and even weaponized particle colliders. The result is that colonizing our solar system, galaxy, and beyond will very likely yield an anarchic state whereby all actors are perpetually in fear of being destroyed when they aren't actively engaged in devastating wars with each other (Torres 2018c; Deudney forthcoming).

Suffice it to say, for the present purposes of this paper, that there are at least some strong *prima facie* reasons for rejecting the "common wisdom" that space colonization constitutes something like an "existential panacea." It could, to the contrary, make our existential situation even more hazardous.

(iv) Humans are bad at estimating the probability of conjunctive and disjunctive propositions. For example, many believe that details increase the probability of stories, whereas just the opposite is true (this is the "conjunction fallacy"). Similarly, many fail to grasp that the more disjuncts are added to a proposition, the more probable the proposition (this is the "disjunctive fallacy"). Thus, "X or Y" is more probable than "X" or "Y," just as "X or Y or Z or A or B or C" is more probable than "X or Y." The point is that the outcomes of extinction or a permanent loss of potential are *causally disjunctive*. They could occur as a result of many disaster scenarios, from catastrophic climate change and asteroid impacts to global nuclear war and poorly implemented stratospheric geoengineering. The more scenarios there are, the greater the overall probability of disaster—and indeed, as mentioned in section 1, the total number of scenarios has been steadily growing since the middle of last century. (This is one advantage of the etiological approach: it maps out the many causal routes to disaster.) Thus, the disjunctive fallacy could lead some to underestimate the likelihood of an existential catastrophe. The same goes for the *observation selection effect*. This arises from the fact that some catastrophes are incompatible with the existence of observers like us. It follows that, independent of the actual probability of extinction per unit of time, we will only ever find ourselves in a world in which no extinction events lie in our evolutionary past. As Milan Ćirković summarizes the idea, "people often erroneously claim that we should not worry too much about existential disasters, since none has happened in the last thousand or even million years. This fallacy needs to be dispelled" (Ćirković 2008).

Other phenomena that could lead people to underestimate or dismiss the threat of annihilation include the confirmation bias, availability bias, gambler's fallacy, affect heuristic,

anchoring, base rate fallacy, optimism bias, overconfidence (Yudkowsky 2008), good-story bias (Bostrom 2002), brain lag (Williams 2002), and what Günther Anders calls "apocalyptic blindness." This

determines a notion of time and future that renders human beings incapable of facing the possibility of a bad end to their history. The belief in progress, persistently ingrained since the Industrial Revolution, causes the incapability of humans to understand that their existence is threatened, and that this could lead to the end of their history (Ehgartner et al. 2017).

(v) There are also reasons for expecting some catastrophes to cluster together in time, a phenomenon called "catastrophe clustering." There are at least three reasons for this. First, a probability theoretic reason: random events will tend to cluster together, thus yielding the "clustering illusion." The question is, therefore, whether we have reason to expect catastrophes to be random. It turns out that we do: on the one hand, many natural phenomena, such as asteroid impacts, appear to be random; on the other hand, studies show that even anthropogenic events like wars, both large and small, have occurred exactly as one would predict if their onset and termination were randomly timed (see Richardson 1960). This gives us reason for thinking that anthropogenic catastrophes of the sort discussed in section 2.2 could also occur in random fashion, thus clustering together in time.

Second, insofar as destroying the world *entirely* is easier if the world has been destroyed *partially*, then terror agents may opportunistically initiate a global attack following a previous disaster. By way of an example, consider that "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, swing flu, Ebola, and Zika" (Senthilingam 2017). If another global outbreak were to occur—perhaps one with consequences comparable to the 1918 Spanish flu; after all, epidemic fatalities on a log scale "tend to follow a power law with a small exponent" (Pamlin and Armstrong 2015)—the result would be panic, confusion, and resource transfer to combat the infection. This could provide an excellent opportunity for malicious agents who have avoided illness to strike a devastating blow to civilization that could have synergistic, rather than additive, effects. Consider also that there have been periods, e.g., in the US and Europe between 1970 and 1993, when terrorist attacks triggered subsequent attacks (Jenkins et al. 2016), and significant evidence supports the hypothesis that mass killings increase the probability (for 13 days on average after the initial event) of copycat shootings (Towers et al. 2015).

Third, Seth Baum and his colleagues outline 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, war, political power shift, economic recession, and so on. Suddenly terminating a SRM project, especially one involving stratospheric geoengineering with sulfate aerosols, could wreak havoc on the global climate, bringing about global agricultural failures or initiating a runaway greenhouse effect (Baum et al. 2013).

(vi) Another problem with nontrivial implications concerns the potential effects of growing ambient CO2 concentrations on human cognition. Not only are the carbon emissions of civilization forcing global climatic changes, but preliminary research suggests that they are quite literally making humans "dumber" (Grossman 2016). Studies show that carbon

dioxide concentrations can have appreciable negative effects on cognition. One study, for example, found "moderate" declines in cognitive performance on decision tasks when the concentration of carbon dioxide was 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). By comparison, our ancestors spent some ~200,000 years breathing in air with between 180 and 280 ppm of CO2. As mentioned in subsection 2.1, we recently passed the milestone of 400 ppm of carbon dioxide in the ambient atmosphere, which is irreversible in the foreseeable future, and carbon dioxide levels could reach upwards of 1,000 ppm by the end of this century (IPCC 2018). Consequently, there could be widespread cumulative effects on our capacity to solve problems at *precisely the moment* when we will confront problems of unprecedented magnitude, complexity, and importance.

But CO2 isn't the only danger to our intellectual vivacity. In the last 40 years or so, over 20,000 new chemicals have been introduced to the market prior to being tested. The fact is that contemporary humans are exposed to a staggering number of potentially toxic—and neurotoxic—agents. Indeed, David Bellinger argues that "Americans have collectively forfeited" 41 million IQ points "as a result of exposure to lead, mercury, and organophosphate pesticides" (see Hamblin 2014). Other chemicals that pose threats to the brain are arsenic, toluene, DDT/DDE, tetrachloroethylene, cadmium, PBDEs, methanol, ethanol, acrylamide, chlorpyrifos, manganese, PCBs, BPA, fluoride, and, perhaps, the cocktail of prescription drugs found in public drinking water (see Boerner 2014). Some of these are common, including BPA in thermal paper receipts and fluoride in PCBs in high-fat foods. Yet the dangers to brain health are far more ubiquitous than this list itself implies. Studies have linked phenomena like 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 jet lag to cognitive impairment. The result is what Christopher Williams calls "environmentally-mediated intellectual decline" (EMID), which has positive and negative manifestations: the former occurs when, e.g., heavy metals are present in one's body, whereas the latter occurs when, e.g., an individual suffers from malnutrition (Williams 1997). Sadly, denizens of the developing world are far more susceptible to EMID than those in the developed world, who have exhibited an overall increase in average IQs in the past few decades, although this trend appears to have stopped and may even be reversing in some regions (see, e.g., Pietschnig and Gittler 2015).[12]

From a societal perspective, even a small, subclinical diminution of average IQ can have significant effects. As Bostrom (2008) observes, a nootropic drug that improves cognitive performance by only 1 percent "would hardy 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 very bad outcome with respect to our collective capacity to solve the global problems before us.

(vii) A related issue concerns research showing that the degree to which people discount

[12]In fact, one explanation for the Flynn effect is the reduction of lead exposure.

the future is, in part, a function of how stable their environments are. When one's environment is unstable, people tend to discount the future more. The reason is, as Pinker (2011) writes, "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." Thus, we should expect that the "context risk" of environmental degradation will lower interest among the public, political leaders, and even scientists in existential risks, given that society will be increasingly preoccupied with more immediate concerns like extreme weather, megadroughts, coastal flooding, sea-level rise, aridification, food supply disruptions, disease outbreaks, mass migrations, social upheaval, economic uncertainty, and the other phenomena listed in section 2.2. The correlation between environmental vagaries and steeper discounting rates is bad news for existential risk scholarship—and, therefore, bad news for human survival.

(viii) Toxic masculinity also poses a major threat to humanity. Consider that the overwhelming number of interstate and civil wars, terrorist attacks, rampage shootings, serial killings, homicides, assaults, rapes, cases of domestic violence and cruelty to animals, and hate crimes are perpetrated by men.[13] (The only category of crime that women consistently have higher arrest rates for is prostitution.) Carl Sagan, following Alan Alda, warns about the prevalence of "testosterone poisoning," which can lead to aggression and violence. Similarly, David Pearce (2012) 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 nonhumans. Directly or indirectly, this century we are likely to kill many more. But perhaps we'll do so in more sophisticated ways.

It is considerations like these that lead Persson and Savulescu (2011) to argue for moral bioenhancement interventions that target men in particular. 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." I have elsewhere criticized Persson and Savulescu's moral bioenhancement thesis (Torres 2017a); nonetheless, such criticisms don't detract from the important point that women, who are consistently underrepresented in decisions to start wars and make peace, should play a much larger role in shaping the developmental trajectory of civilization. In fact, one study suggests that the *only* variable that is directly and positively correlated with the "collective intelligence" of groups (analogous to psychometric g in individuals) is the number of women within the group—i.e., the more women, the smarter the collective (Woolley et al. 2010). It follows that insofar as mitigating existential risk is a group activity, the community should want more women scholars.

In sum, a male-dominated geopolitical arena in which unprecedentedly powerful tech-

[13]Not to mention that male psychopaths outnumber female psychopaths by 20 to 1.

nologies are increasingly accessible to state and nonstate actors could greatly increase the overall probability of doom. Toxic masculinity, testosterone poisoning, and the statistical lack of altruism among men pose serious threats to our survival.

(ix) Finally, we have so far primarily examined risks to human survival from an *a posteriori* perspective. But there are also *a priori* issues that should be included in any evaluation of the human existential predicament. For example, consider that the two most plausible principles for generating self-locating beliefs both appear to support a pessimistic outlook. First, the "self-sampling assumption" (SSA) leads to the Doomsday Argument, which concludes that we are systematically underestimating the probability of doom. That is, whatever our prior estimates of annihilation are, we should increase them (see Leslie 1996).[14] The second, alternative principle is the "self-indication assumption" (SSA). This dodges the Doomsday Argument, but within the well-established Great Filter framework it yields perhaps an even more worrisome conclusion, namely, that the end is near, i.e., there exists a Great Filter between our current stage of technological development and the next stage (Grace 2010; Hanson 2010). Thus, whether one adopts SIA or SSA, *the prospects for human survival appear dimmer than empirical analyses alone would suggest.*

4. Conclusion

As I have noted elsewhere, our planetary spaceship, Earth, will remain habitable for another 1 billion years, or 10 million centuries. To put this in perspective, our species, the self-described "wise man," has roamed this oblate spheroid, stuck between the dusty ground and the infinite firmament, for a mere 2,000 centuries. If we consider the emergence of recorded history about 4,000 years ago to be the beginning of civilization, then we may have progressed through a mere 0.0002 percent of our entire potential history on Earth. If we were to map this onto the annual calendar—*a la* the Cosmic Calendar—then we are no more than about 63 seconds into the first day of the year. Following the Long Now Foundation's decision to write the current year as "02018" to encourage "deep time" thinking, we can write the year as "0000002018" to emphasize the potential habitability of our planet. Even more, some scholars have estimated that a *million billion human beings* could occupy Earth before the oceans evaporate and the sun sterilizes the planet, assuming currently normal lifespans and a population of 1 billion or more (Bostrom 2013). There could be orders of magnitude more if we successfully colonize our supercluster, or beyond. Thus, the stakes are incredibly high; there is a lot to lose by succumbing to one of the Great Challenges. Yet a Google Scholar search reveals a mere 1,910 results for the word "existential risk," compared to 2,060 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 dwarfed by 5,390,000 results for "cancer."[15] The unfortunate reality is that existential risks in general are severely understudied.

The aim of this paper is to offer a comprehensive map of the obstacle course of risks before us, focusing on the three Great Challenges of climate change and the Anthropocene

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

[15]These results were recorded on January 24, 2018, around 8:30 pm. The number of results for "suffering risks" was far less than any number here presented.

extinction, the distribution of unprecedented offensive capabilities across society, and value-misaligned machine superintelligence. In doing so, I hope to have convinced readers that the dangers are real and urgent. The importance of recognizing this is further underlined by the fact that *there is not a single threat on the road ahead that is fundamentally insoluble*. With concerted efforts, humanity can easily slalom around the biggest dangers. The first step toward accomplishing this feat, though, is to understand exactly what we, the self-described "wise man," are up against.[16]

Appendix 1: A typology of risks organized by their causes (see Torres 2017a).

References:

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.

Allison, Graham. 2004a. Nuclear Terrorism: The Ultimate Preventable Catastrophe. New York, NY: Owl Books.

Allison, Graham. 2004b. How to Stop Nuclear Terror, Foreign Affairs. https://www.foreignaffairs.com 01-01/how-stop-nuclear-terror.

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, and Kaj Sotala. 2012. How We're Predicting AI—or Failing To. Machine Intelligence Research Institute. https://pdfs.semanticscholar.org/5a12/80f783e4ce6ba31b821f4d86f

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. 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.

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.

[16]This article draws heavily from numerous articles of mine, including Torres 2016 and 2017. In some cases, strings of sentences are reproduced ad verbum, while in others the same ideas are paraphrased. Please see these more detailed documents for additional information about existential risks.

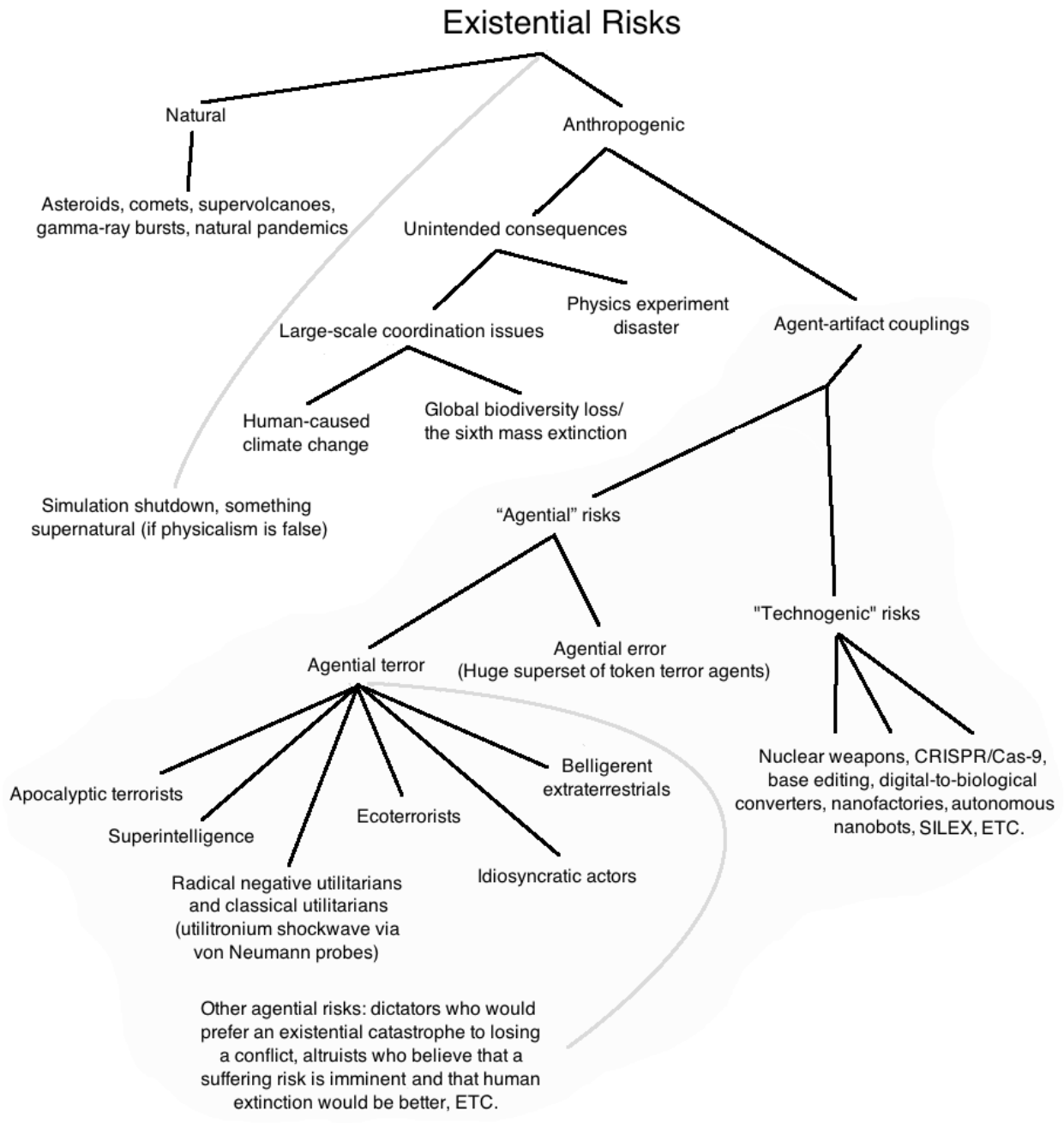
Existential Risks
Natural
Anthropogenic
Asteroids, comets, supervolcanoes, gamma-ray bursts, natural pandemics
Unintended consequences
Physics experiment disaster
Large-scale coordination issues
Agent-artifact couplings
Human-caused climate change
Global biodiversity loss/ the sixth mass extinction
Simulation shutdown, something supernatural (if physicalism is false)
"Agential" risks
"Technogenic" risks
Agential terror
Agential error (Huge superset of token terror agents)
Apocalyptic terrorists
Superintelligence
Ecoterrorists
Belligerent extraterrestrials
Idiosyncratic actors
Radical negative utilitarians and classical utilitarians (utilitronium shockwave via von Neumann probes)
Nuclear weapons, CRISPR/Cas-9, base editing, digital-to-biological converters, nanofactories, autonomous nanobots, SILEX, ETC.
Other agential risks: dictators who would prefer an existential catastrophe to losing a conflict, altruists who believe that a suffering risk is imminent and that human extinction would be better, ETC.

Figure 4:

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

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

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.

Brin, David. 1983. The "Great Silence"—The Controversy Concerning Extraterrestrial Intelligent Life. Quarterly Journal of the Royal Astronomical Society. 24(3): 283-309.

Brin, D. 1998 The Transparent Society: Will Technology Force Us To Choose Between Privacy And Freedom?, Perseus Books Group.

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/.

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.o

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

Chomsky, Noam. 2010. Human Intelligence and the Environment. https://chomsky.info/20100930/.

Cirincione, J. 2015 Nuclear Nightmares: Securing the World Before It Is Too Late, Columbia University Press.

Církovic, Milan. 2008. Observation Selection Effects and Global Catastrophic Risks. In Nick Bostrom and Milan Církovic (eds.), Global Catastrophic Risks. Oxford, UK: Oxford University Press.

Cohen, N. Health and the Rise of Civilization. New Haven, CT: Yale University Press.

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

Cotton-Barratt, Owen, and Toby Ord. 2015. Existential Risk and Existential Hope: Definitions. Future of Humanity Technical Report #2015-1. https://www.fhi.ox.ac.uk/Existential-

risk-and-existential-hope.pdf.

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.

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).

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/.

Farber, D. 2010 Nuclear Attack a Ticking Time Bomb, Experts Warn, CBS News. https://www.cbsnews.com/news/nuclear-attack-a-ticking-time-bomb-experts-warn/.

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.

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.

Gloor, Lukas, and Adriano Mannino. 2018. The Case for Suffering-Focused Ethics. Foundational Research Institute. https://foundational-research.org/the-case-for-suffering-focused-ethics/.

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
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/170

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. 1998 The Great Filter—Are We Almost Past It? Unpublished Manuscript. http://mason.gmu.edu/˜rhanson/greatfilter.html.

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.

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/
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/ar

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_char

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, Kevin. 2008. The Expansion of Ignorance. The Technium. URL: http://kk.org/thetechnium/th expansion-o/.

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.

Koebler, Jason. 2016. Society Is Too Complicated to Have a President, Complex Mathematics Suggest. Motherboard. https://motherboard.vice.com/en_us/article/wnxbm5/society-is-too-complicated-to-have-a-president-complex-mathematics-suggest.

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.

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.

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_Pantuc

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.

Pearce, David. 2012. Transhumanism and the Abolitionist Project. City Magazine. https://www.hedweb.com/transhumanism/2012-interview.html.

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.

PEW, 2015 The Future of World Religions: Population Growth Projections, 2010-2050. http://www.pewforum.org/2015/04/02/religious-projections-2010-2050/.

Phoenix, C. 2008 Estimating a Timeline for Molecular Manufacturing, Center for Responsible Nanotechnology. http://www.crnano.org/timeline.htm.

Pietschnig, Jakob, and Georg Gittler. 2015. A Reversal of the Flynn Effect for Spatial Perception in German-Speaking Countries: Evidence from a Cross-Temporal IRT-Based Meta-Analysis (1977-2014). Intelligence. 53: 145-153.

Pinker, S. 2011 The Better Angels of Our Nature: Why Violence Has Declined, 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://abcnews.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.

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/
13-17.pdf

Russell, Bertrand. 1954. Human Society in Ethics and Politics. New York, NY: George Allen & Unwin Ltd.

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

Sandberg, Anders. 2014. The Five Biggest Threats to Human Existence. The Conversation. https://theconversation.com/the-five-biggest-threats-to-human-existence-27053.

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

Sandberg, A., Bostrom, N. 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.

Samuels, Brett. 2017. UN Secretary-General Issues "Red Alert" for World Ahead of 2018. The Hill. http://thehill.com/policy/international/366940-un-secretary-general-issues-red-alert-for-world-ahead-of-2018.

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=.

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

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.

Smallman, Elton. 2016. Climate Change Specialist Predicts Human Extinction in 10 Years. Stuff. https://www.stuff.co.nz/environment/86778981/climate-change-specialist-predicts-human-extinction-in-10-years.

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/5nC treasury.gov.uk/sternreview_index.htm.

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

Taleb, N. 2016. The "Long Peace" Is a Statistical Illusion. http://www.fooledbyrandomness.com/pinke

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

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

Tonn, B., MacGreagor, D., 2009 Are We Doomed? Futures 41(10): 673-675.

Torres, Phil. 2016a. It Matters Which Trend Lines One Follows: Why Terrorism Is an Existential Threat. *Free Inquiry*. http://media.wix.com/ugd/d9aaad_3b1dd6abb92744dd8474ff1c57ff676

Torres, Phil. 2016b. *The End: What Science and Religion Tell Us About the Apocalypse*. Durham, NC: Pitchstone Publishing.

Torres, Phil. 2016c. Existential Risks Are More Likely to Kill You Than Terrorism. Future of Life Institute. http://futureoflife.org/2016/06/29/existential-risks-likely-kill-terrorism/.

Torres, Phil. 2017a. *Morality, Foresight, and Human Flourishing: An Introduction to Existential Risks*. Durham, NC: Pitchstone Publishing.

Torres, Phil. 2017b. Agential Risks and Information Hazards: An Unavoidable but Dangerous Topic? *Futures*. https://docs.wixstatic.com/ugd/d9aaad_d9c1dceb74df4dccab498e50d296746d.pd

Torres, Phil. 2017c. Three Minutes Before Midnight: An Interview with Lawrence Krauss About the Future of Humanity. *Free Inquiry*. https://www.secularhumanism.org/index.php/artic

Torres, Phil. 2018a. Who Would Destroy the World? Omnicidal Agents and Related Phenomena. *Aggression and Violent Behavior*. (forthcoming) https://docs.wixstatic.com/ugd/d9aaad_e

Torres, Phil. 2018b. Superintelligence and the Future of Governance: On Prioritizing the Control Problem at the End of History. In Roman Yampolskiy (ed.), *Artificial Intelligence Safety and Security*. New York: Taylor and Francis Group. (forthcoming) https://docs.wixstatic.com/ugd/d9aaad_34d10a04399e4547978bb834d65cbcba.pdf.

Torres, Phil. 2018c. Space Colonization and Suffering Risks: Reassessing the Maxipok Rule. Forthcoming. https://goo.gl/j79ipg.

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

Turchin, A., Green, B.R. 2017 Aquatic Refuges for Surviving a Global Catastrophe, Futures 89: 26-37.

UN. 2017. World Population Prospects. https://esa.un.org/unpd/wpp/Publications/Files/WPP2017

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.

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

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.

Wittes, B., Blum, G. 2015 The Future of Violence: Robots and Germs, Hackers and Drones—Confronting A New Age of Threat, Basic Books.

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_repor

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

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