

2019

Long-Term Trajectories of Human Civilization

Seth D. Baum

Global Catastrophic Risk Institute

Stuart Armstrong

University of Oxford

Timoteus Ekenstedt

Umeå University

Olle Häggström

Chalmers University of Technology

Robin Hanson

*George Mason University**See next page for additional authors*Follow this and additional works at: https://scholarworks.smith.edu/eco_facpubsPart of the [Economics Commons](#)

Recommended Citation

Baum, Seth D.; Armstrong, Stuart; Ekenstedt, Timoteus; Häggström, Olle; Hanson, Robin; Kuhlemann, Karin; Maas, Matthijs M.; Miller, James D.; Salmela, Markus; Sandberg, Anders; Sotala, Kaj; Torres, Phil; Turchin, Alexey; and Yampolskiy, Roman V., "Long-Term Trajectories of Human Civilization" (2019).

Economics: Faculty Publications, Smith College, Northampton, MA.

https://scholarworks.smith.edu/eco_facpubs/27

This Article has been accepted for inclusion in Economics: Faculty Publications by an authorized administrator of Smith ScholarWorks. For more information, please contact scholarworks@smith.edu

Authors

Seth D. Baum, Stuart Armstrong, Timoteus Ekenstedt, Olle Häggström, Robin Hanson, Karin Kuhlemann, Matthijs M. Maas, James D. Miller, Markus Salmela, Anders Sandberg, Kaj Sotala, Phil Torres, Alexey Turchin, and Roman V. Yampolskiy

Long-Term Trajectories of Human Civilization

Published in *Foresight* 21(1):53-83 (2019), [DOI 10.1108/FS-04-2018-0037](https://doi.org/10.1108/FS-04-2018-0037).

This version 11 March 2019.

Seth D. Baum,¹ Stuart Armstrong,² Timoteus Ekenstedt,³ Olle Häggström,⁴ Robin Hanson,⁵ Karin Kuhlemann,⁶ Matthijs M. Maas,⁷ James D. Miller,⁸ Markus Salmela,⁹ Anders Sandberg,² Kaj Sotala,¹⁰ Phil Torres,¹¹ Alexey Turchin,¹² and Roman V. Yampolskiy¹³

1. Global Catastrophic Risk Institute, USA. Corresponding author: seth@gcrinstitute.org
2. Future of Humanity Institute, University of Oxford, UK
3. Umeå University, Sweden
4. Department of Mathematical Statistics, Chalmers University of Technology, Sweden
5. Department of Economics, George Mason University, USA
6. Department of Political Science, University College London, UK
7. Centre for International Law, Conflict and Crisis, University of Copenhagen, Denmark
8. Department of Economics, Smith College, USA
9. Independent researcher, Sweden
10. Foundational Research Institute, Germany
11. Independent researcher, Canada
12. Science for Life Extension Foundation, Russia
13. Department of Computer Engineering and Computer Science, University of Louisville, USA

Abstract

Purpose: This paper formalizes long-term trajectories of human civilization as a scientific and ethical field of study. The long-term trajectory of human civilization can be defined as the path that human civilization takes during the entire future time period in which human civilization could continue to exist.

Approach: We focus on four types of trajectories: *status quo trajectories*, in which human civilization persists in a state broadly similar to its current state into the distant future; *catastrophe trajectories*, in which one or more events cause significant harm to human civilization; *technological transformation trajectories*, in which radical technological breakthroughs put human civilization on a fundamentally different course; and *astronomical trajectories*, in which human civilization expands beyond its home planet and into the accessible portions of the cosmos.

Findings: Status quo trajectories appear unlikely to persist into the distant future, especially in light of long-term astronomical processes. Several catastrophe, technological transformation, and astronomical trajectories appear possible.

Value: Some current actions may be able to affect the long-term trajectory. Whether these actions should be pursued depends on a mix of empirical and ethical factors. For some ethical frameworks, these actions may be especially important to pursue.

Keywords: long-term trajectories, human civilization

1. Introduction

What will human civilization look like in one million, one billion, or one trillion years? These are questions of broad scientific and ethical significance, yet they are neither well-studied nor well-understood. While cosmology has probed stellar and physical dynamics

in the deep future (Adams 2008), studies of human futures in demography, economics, sustainability science, political science, and related disciplines tend to concentrate on upcoming decades, often relying on simple extrapolations of existing trends, or on scenario writing. But important civilizational processes could play out over longer time scales. To restrict attention to near-term decades may be akin to the drunk searching for his keys under the streetlight: it may be where empirical study is more robust, but the important part lies elsewhere.

In this paper, we seek to formalize long-term trajectories of human civilization as a scientific and ethical field of study. We synthesize perspectives from a range of fields, including moral philosophy, demography, economics, sustainability science, risk analysis, futures studies, political science, archaeology, climatology, and astrobiology. Using insights from these fields, we establish four broad classes of trajectories and describe important details for each. The four classes of trajectories are:

- (1) *Status quo trajectories*, in which human civilization persists in a state broadly similar to its current state into the distant future
- (2) *Catastrophe trajectories*, in which one or more events cause significant harm to human civilization
- (3) *Technological transformation trajectories*, in which radical technological breakthroughs put human civilization on a fundamentally different course
- (4) *Astronomical trajectories*, in which human civilization expands beyond its home planet and into the accessible portions of the cosmos

These four classes of trajectories represent the major potential long-term trajectories of human civilization. They depict human civilization staying roughly level (1), shrinking (2), or potentially expanding via technological breakthrough (3) or astronomical expansion (4). Scenarios that do not fit in any one of these four classes appear unlikely. Some scenarios can fit in multiple classes: for example, technological breakthrough could lead to either catastrophe or astronomical expansion. However, each class contains distinct features worthy of separate consideration and drawing on separate scholarly traditions. Status quo trajectories draw from the trend extrapolations of demography, economics, sustainability science, and related disciplines (e.g., Field et al. 2014; United Nations 2017a). Catastrophe trajectories draw from analysis of global catastrophic and existential risks, in particular studies of the aftermath of these events (e.g., Maher and Baum 2013; Dartnell 2014; Denkenberger and Pearce 2014). Technological transformation trajectories draw from futures studies, especially related to technological singularity and related notions of radical technological change (e.g., Miller 2012; Eden et al. 2012; Hanson 2016). Finally, astronomical trajectories draw from astrobiology, relating the socio-technological possibilities of space travel to the astronomical structure of the universe (e.g., Ćirković 2002; Haqq-Misra and Baum 2009; Armstrong and Sandberg 2013). To our knowledge, these various lines of inquiry have not previously been synthesized into a formal study of long-term trajectories of human civilization.¹

¹ Some prior studies covering similar ground include Maher and Baum (2013), which analyzes long-term trajectories mainly in the context of catastrophe scenarios but with some attention to status quo and astronomical trajectories, and Beckstead (2013), which develops moral philosophy arguments for caring about long-term trajectories.

The study of long-term trajectories is further important because some current actions may affect long-term outcomes. These include actions that affect the risk of catastrophe, the onset of transformative technologies, or the expansion of human civilization into outer space. How these actions should be evaluated depends on empirical details about the nature of long-term trajectories as well as ethical theories concerning the trajectories' valuation.

In Section 2, we further develop the conceptual and ethical foundations of long-term trajectories of human civilization. We argue that the trajectories can be defined and evaluated in a variety of ways, consistent with a variety of empirical and ethical perspectives. We discuss details of status quo and catastrophe trajectories in Sections 3 and 4. Section 5 presents sketches of possible technological transformation and astronomical trajectories. We discuss details of technological transformation and astronomical trajectories in Sections 6 and 7. Section 8 concludes.

2. Conceptual and Ethical Foundations

The long-term trajectory of human civilization can be defined as the path that human civilization takes into the long-term future. The long-term can in turn be defined as the entire future time period in which human civilization could continue to exist. Given uncertainty about which trajectory human civilization will take, we speak in terms of trajectories in the plural, to emphasize the variety of trajectories that appear possible from our current vantage point.

Human civilization is a more complex concept. Dictionary definitions of civilization emphasize an advanced state of cultural, organizational, social, and technological development.² “Human” could be defined as the species *Homo sapiens sapiens*. However, over long time scales, human descendants are likely to become a different species via genetic drift (Wills 2008). It may also be possible for humans (or their descendants) to engineer new biological species or non-biological beings (e.g., robots) that are capable of continuing civilization into the future (More and Vita-More 2010). Therefore, this paper uses the term “human civilization” to refer to any civilization that traces to the current human population. This definition includes civilizations led by genetic descendants of *Homo sapiens sapiens*, as well as civilizations led by biological or non-biological beings that are engineered by *Homo sapiens sapiens* or its genetic descendants. This definition permits a study of long-term trajectories that does not need to constantly account for whether the civilization is still in some sense “human”.

It will often be advantageous to think of the trajectories in quantitative terms, in the form of one or more time series. In other words, specific trajectories are what one obtains from plotting one or more key attributes of human civilization (on which more shortly) as a function of time, with time going from the present into the long-term future. A full understanding of long-term trajectories also requires qualitative description of what is going on in the trajectories: the story behind the numbers.

Figure 1 shows some illustrative trajectories. These trajectories are for illustrative purposes only; no precision is intended. Also, these are not the only forms that the trajectories could take. Indeed, as discussed throughout this paper, the trajectories could have a variety of forms and could in some cases even be negative. Note that Figure 1

² See e.g. <https://www.merriam-webster.com/dictionary/civilization>;
<https://en.oxforddictionaries.com/definition/civilization>.

does not label the vertical axis. Instead, it uses an unspecified aggregate measure of human civilization. Other figures throughout this paper do the same. This is done in deference to the possibility of using a variety of measures for the total size of a civilization. Plots of attributes other than total size, such as averages across the population or the degree of inequality within the population, may appear differently.

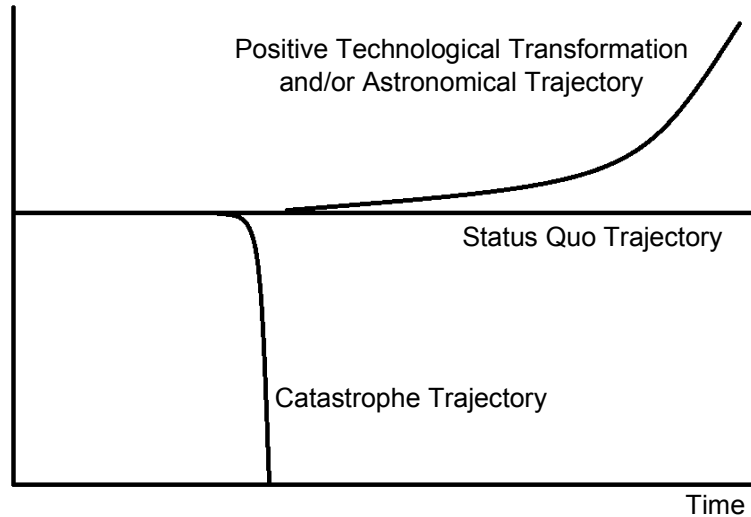


Figure 1. Illustrative trajectories.

This raises the question of which attributes of human civilization should be focused on. Studies of near-term trajectories tend to focus on population, economic production, quality of life, and select natural resources and other environmental parameters. These attributes can also be considered for long-term trajectories, though with some complications. Over the long-term, human civilization may not be using the same natural resources as it currently is, and the environment may change sufficiently that near-term environmental parameters are unimportant. The form of civilization may likewise change enough that current conceptions of economic production, security, and quality of life do not meaningfully apply. Even current conceptions of population could be inapplicable to some long-term scenarios, such as those in which civilization is led by something other than *Homo sapiens sapiens*.

Given the wide range of long-term possibilities, we call for greater care in the selection of attributes for quantifying long-term trajectories than for short-term trajectories, and for greater emphasis on qualitative descriptions to clarify what is going on in different trajectories.

As with near-term trajectories, the selection of attributes can also depend on ethical factors. While a complete review of ethical theories is beyond the scope of this paper, what follows is a partial overview.³ Some ethical theories value total quality of life aggregated across the population, while others value average quality of life of members of the population, and still others favor improvements for the less well-off members of the population (Ng 1989; Arrhenius 2000; Adler 2012). The average-based theories

³ More detailed discussions of ethical issues of relevance to long-term trajectories and other future outcomes can be found in, among other sources, Portney and Weyant (1999), Weisbach and Sunstein (2007), Schwartz and Milligan (2016), Scheffler (2018), and Tonn (2018).

would have less interest in tracking the population in different trajectories, while those focused on the less well-off members would be especially interested in inequality in distributions across population members for different trajectories. A different type of theory values the survival and flourishing of humanity into the future, but does not value its instantaneous size or per-person metrics (Schwartz 2011; Frick 2017). Ethical theories also diverge on the (dis)value of deception, i.e. on whether ignorance is bliss (Nozick 1974; Ng 1990; Sagan 2006).⁴ These theories point to different attributes, especially for scenarios involving large-scale deception, such as if human minds are confined to virtual reality environments.⁵ Other points of divergence are on whether the natural environment (Norton 1984; Agar 2000; Morito 2003) or non-human animal welfare (Johansson-Stenman 2018) are of intrinsic value—valuable for its own sake—or only of instrumental value as something that improves human civilization. This divergence points in different directions for scenarios in which human civilization thrives while causing greater and greater harm to natural environments or non-human animal populations. Finally, there is also divergence on how much importance to place on suffering and/or other negative value relative to happiness and/or other positive value (e.g., Baum 2008; Gloor and Mannino 2016; Sotala and Gloor 2017). Ethical theories emphasizing negative value would tend to be more interested in tracking the long-term trajectories of negative value, and vice versa for positive value. Indeed, ethical theories emphasizing negative value could even consider the survival and growth of the human population to be a bad thing if it substantially increased the amount of suffering.

These are among the reasons why ethical theory is an important part of the study of long-term trajectories. Another important reason is for the relation of long-term trajectories to contemporary human affairs. This is because some current actions could affect the long-term trajectory of human civilization, including actions that change the risk of major catastrophes, the prospects for radical technological breakthrough, and/or the prospects for the colonization of outer space. The importance of these actions depends on their effect on long-term trajectories and on how much long-term trajectories are valued. Some ethical theories have inter-temporal neutrality, valuing long-term effects as much as short-term effects, while other ethical theories discount future values or only value that which affects existing persons (Parfit 1984; Arrhenius 2003; Weinberg 2008).⁶ Interpreting the significance of long-term trajectories for contemporary human affairs requires attention to these and other ethical issues.

As an illustration of the significance of long-term trajectories for contemporary human affairs, consider the case of time-neutral total utilitarianism, in which all welfare—human or otherwise—is valued equally, regardless of when it occurs. We stress that

⁴ Nozick (1974: 44-45) argues against “experience machines” that provide people with seemingly fulfilling but false virtual realities; Sagan (2006: 217-218) argues against ignorance and proposes an 11th commandment of “Thou shalt understand the world, figure things out”. In contrast, Ng (1990) argues that accurate knowledge should only be valued to the extent that it increases welfare, and it should be disvalued if it reduces welfare.

⁵ For one argument to this effect, see Bostrom (2003). Such scenarios are dramatized in the popular film *The Matrix*. An alternative perspective on these scenarios proposes that they are not actually deceptive, but instead involve different processes underlying the same reality we believe to exist (Chalmers 2005).

⁶ Ethical theories that only value existing persons may nonetheless place some concern on long-term trajectories insofar as existing people care about long-term trajectories; see Finneron-Burns (2017, Section 2.4).

this view is selected for illustration purposes only and is not intended as an endorsement,⁷ nor do we claim that this is the only view for which long-term trajectories are important. This view would place great value on long-term trajectories involving large amounts of welfare. This leads to some potential claims about what current decision making should prioritize:

- (A) Reducing the risk of human extinction, because extinction would result in the loss of all future generations.
- (B) Reducing the risk of major catastrophes that would lead to the permanent loss of advanced human civilization.
- (C) Expediting technological breakthroughs and ensuring that they would improve welfare.
- (D) Expediting space colonization and ensuring that it would improve welfare.
- (E) Improving near-term welfare, because long-term trajectories will be the same either way, or because the effect of current decisions on long-term trajectories is not understood well enough to guide decision making.

One can imagine a case for each of these five claims. (A) could hold if humanity faces extinction threats and if extinction is the only thing that could change its long-term trajectory. (B) could hold if humanity faces major catastrophic risks and if such catastrophes would diminish the long-term trajectory. (C) could hold if technological breakthroughs can be achieved before catastrophe occurs. (D) could hold if space colonization can be achieved before catastrophe occurs and before a technology breakthrough occurs. Finally, (E) could hold if contemporary actions would not affect long-term trajectories.

As noted in Section 1, attention typically goes to (E), though without consideration of long-term trajectories. Prior literature studying the matter has tended to support either (A) or (B). Sagan (1983), Parfit (1984), Ng (1991), and Matheny (2007) argue for prioritizing the reduction of human extinction risk on grounds that extinction would end all future generations. Bostrom (2013) and Maher and Baum (2013) argue for prioritizing the reduction of the risk of any catastrophe that would cause significant long-term harm to human civilization. Similarly, Beckstead (2013) argues for any action that improves long-term trajectories and suggests inconclusively that reducing the risk of these long-term catastrophes may be the most effective approach. However, this prior literature generally lacks careful empirical study of potential long-term trajectories, making it difficult to assess which claim is correct or how to weigh them in decision making.

3. Status Quo Trajectories

Status quo trajectories involve the current civilization continuing in something like its current form into the long-term future. Status quo trajectories do not necessarily involve total stasis—indeed, stasis may be unlikely—but they do involve the continuation of current trends without major discontinuities.

Exactly what qualifies as a status quo trajectory is a fuzzy and debatable matter. It could refer to continuity in the state of human civilization, or continuity in the trends of

⁷ Indeed, some of the authors of this paper disagree with this view.

change in the state, or even continuity in trends in changes of trends.⁸ For example, the current human population is steadily increasing at a decreasing rate. Another matter is how large and/or abrupt of a deviation there needs to be for a trajectory to no longer qualify as status quo. The accumulation of status-quo trends in economic and technological development, and in environmental and resource degradation, compounded over decades, centuries, millennia, or even longer, likely results in something very different from the present civilization. It is a matter of interpretation whether this difference is large enough to qualify as outside the status quo. Indeed, some projections of catastrophe derive from extrapolations of current trends, especially trends of global environmental change and natural resource depletion (Rockström et al. 2009; Baum and Handoh 2014), while some projections of radical technological transformation derive from extrapolations of current trends, such as projections of transformative artificial intelligence scenarios based on the “Moore’s law” trend in computer hardware performance (Moravec 1998; Kurzweil 2005). It is similarly a matter of interpretation whether the current civilization is a continuation of the status quos that existed 1,000, 100, or even just 20 or 30 years ago, given the many changes there have been in politics, culture, and technology.

For purposes of this paper, we will restrict status quo trajectories to those with no radical changes to the state of human civilization or its underlying trends. Status quo trajectories would keep important attributes of human civilization in a form that would be broadly recognizable to current observers. (Which attributes of human civilization are important can be derived from ethics and related factors, as discussed in Section 2.) For example, a status quo trajectory could involve significant environmental degradation and technological advancement that combine to keep attributes such as population at a similar size or growth rate. This conception of status quo trajectories permits some changes in the important attributes, but not the more extreme changes that may be possible. Some more extreme changes are discussed in the sections below on the other types of trajectories.

Near-term trajectory analyses often present a range of possible future trajectories, given a range of different assumptions. For example, the United Nations (2017a) offers probabilistic projections of population growth scenarios through 2100, with a 95% prediction interval range of roughly 9.5 billion to around 13.5 billion humans at that time. Figure 2 presents three trajectories of United Nations population data for past and future world populations. Each of these trajectories could be considered status quo.

The above discussion uses metrics based on human attributes. Other ethical perspectives can point in different directions. For example, current aggregate value may be negative if non-human animal welfare is taken into account. The farm animal population is much larger than the human population; an estimated 68 billion non-human animals were slaughtered for food worldwide in 2012 (United Nations 2012). A substantial fraction of them suffer in factory farms (Mallon 2005; DeGrazia 2009), plausibly so much that their existence, and in turn aggregate global value, is net negative. Fortunately, status quo trajectories may change this, such as via “in vitro meat” that could substitute factory farms for painless meat production (Bhat and Fayaz 2011; Post 2012).

⁸ In mathematical terms, this could be the state of human civilization as shown in Figures 2-3 as well as its first and second derivatives with respect to time.

For ethical views that include non-human animal welfare, this is an important development to track.

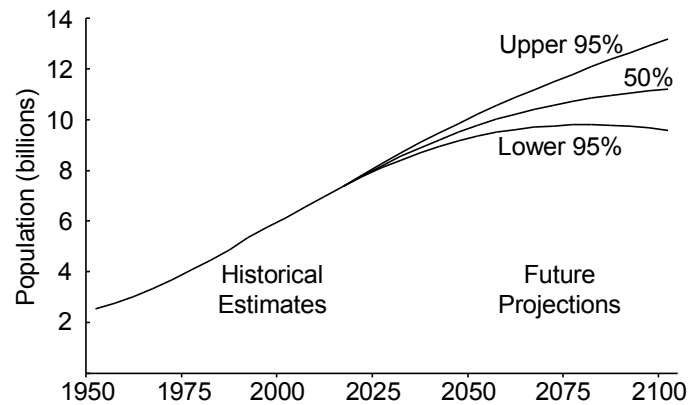


Figure 2. Historical world population estimates for 1950-2015, and future population projections out to 2100 at three prediction intervals, based on data from United Nations (2017b; 2017c).

Over the long-term, the persistence of status quo trajectories appears rather unlikely. Instead, civilization is likely to either die out or transform. If nothing else, civilization would presumably end around a few hundred million to a few billion years from now, when Earth becomes uninhabitable due to the Sun becoming warmer and larger (O'Malley-James et al. 2014; Wolf and Toon 2015). But it would be a remarkable feat for a status quo trajectory to persist for so long. The frequency of major extinction events in Earth's history suggests that the status quo civilization is unlikely to survive for one billion more years. Survival into the far future may require transformative technology and/or astronomical expansion, which may be likely to happen eventually as long as catastrophe is avoided. Indeed, rapid ongoing technological and social change suggests that some sort of major transformation is likely to occur relatively soon. Finally, there have been major changes before, such as from agriculture and industrialization, which further suggests the possibility of future trajectory changes. As noted above, it is a matter of interpretation whether some of these changes push outside the status quo. Still, we find it hard to expect any reasonable interpretation of the status quo to persist into the long-term future.

Perhaps we underestimate the durability of the status quo civilization. Perhaps past upheavals are no longer applicable. Perhaps current civilization is highly resilient to a range of catastrophes and/or agile at avoiding them. Perhaps radical technological breakthroughs and space travel will prove elusive. In that case, the status quo could hold for a long time. But it would seem at least as possible to overestimate the durability of the status quo civilization, given our present vantage point from within it. Our civilization seems quite robust, given its current dominant position on Earth. But this perspective may be biased by the fact that we are within the status quo. Indeed, our status quo world may be less durable and more historically and environmentally contingent than it would seem, including our international relations (e.g., Lebow 2015), our civilization (e.g., Richerson et al. 2001), and our very existence (e.g., Ćirković et al. 2010). Thus, while we cannot rule out the long-term viability of status quo trajectories, other trajectories appear more likely.

4. Catastrophe Trajectories

Among catastrophe trajectories, the simplest to analyze are those involving human extinction. Following extinction, basic attributes including population, economic production, and quality of life all fall to zero.⁹ More complex are catastrophes that some people survive, but in a form that is qualitatively different from the status quo civilization. Analysis of these sub-extinction catastrophes requires attention to the prospects for humans in what could be a radically altered world, and to how readily survivors could rebuild some form of civilization.

Prior research has identified a number of human extinction risks, including nuclear war, collision between Earth and a large asteroid or comet, supervolcano eruption, global warming, runaway artificial intelligence, physics experiment disasters, and scenarios involving multiple major catastrophes (e.g., Asimov 1981; Rees 2003; Leggett 2006; Bostrom and Ćirković 2008; Häggström 2016). Some of these events would likely result in immediate human extinction if they occurred—for example, a physics experiment disaster could alter the astronomical vicinity, rendering life on Earth impossible (Turner and Wilczek 1982; Ord et al. 2010). For other events, the outcome is more ambiguous—for example, asteroids, comets, volcanoes, and nuclear war could all block sunlight, decimating global agriculture, but some people may be able to survive on stored food or food grown from biomass or fossil fuels (Denkenberger and Pearce 2014; Baum et al. 2015a). Tonn and MacGregor (2009) study a “singular chain of events” in which a series of catastrophes befall humanity; they find that, in the absence of a decisive extinction event, it is actually difficult to get to extinction. We find this reasoning compelling while noting the difficulty of assessing unprecedented catastrophe scenarios.

For sub-extinction catastrophes, the long-term trajectories depict the fate of the survivor populations. Will they persist for some extended time in their diminished state, and then die out later on? Will they rebuild something along the lines of the pre-catastrophe, status quo civilization? Will they go on to achieve technological transformation and/or space colonization? The fate of survivor populations depends first on how they fare in the immediate aftermath of the catastrophe and then on their long-term prospects to manage and/or rebuild civilization.

An important variable for catastrophe trajectories is the speed of the catastrophe. Some catastrophes could cause significant harm to human civilization in a short period of time, such as nuclear wars or pandemics. Other catastrophes work more slowly, such as global warming or the depletion of certain natural resources. Slow catastrophes give humans more time to adapt to the new conditions, though the conditions may also be more durable. What follows is written mainly with fast catastrophes in mind, though some of the discussion may also apply to slow catastrophes.

Figure 3 presents a variety of potential catastrophe trajectories. It distinguishes between catastrophes resulting in extinction, in survival with neither agriculture nor industry, and in survival with agriculture but not industry. Agriculture and industry are essential features of modern human civilization and are discussed in detail below. Figure 3 also shows survivor populations either remaining in the post-catastrophe state, recovering back towards the state of the current civilization, or enduring subsequent catastrophes. Finally, Figure 3 shows the current civilization avoiding catastrophe and

⁹ A possible exception would be if automated technology maintains economic production without human upkeep, though current technology is not sufficiently automated for this.

maintaining the status quo trajectory, as well as the current civilization and two post-catastrophe civilizations achieving technological transformation and/or astronomical trajectories. These are all among the possible trajectories a post-catastrophe civilization (or the current civilization if it avoids catastrophe) could take. As with Figure 1, these trajectories are for illustrative purposes only; no precision is intended. Additionally, the vertical axis is not labeled for the same reason as with Figure 1.

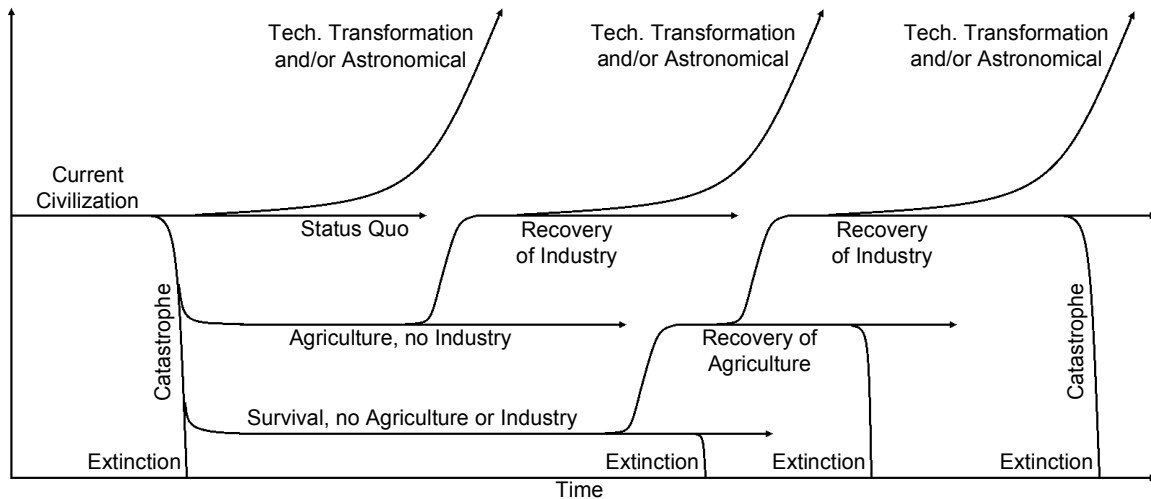


Figure 3. Illustrative catastrophe trajectories.

4.1 The Immediate Aftermath

In the immediate catastrophe aftermath, survivors will need to secure their basic needs, especially food and water. Most of the contemporary human population obtains food and water from civilization: they do not procure their own. This portion of the population could have a relatively difficult time in the immediate aftermath, though this depends on the functioning of civilization. If food and water provision continues, then a relatively large population could survive. Thus, an important question is how resilient the basic functioning of civilization is to major catastrophes.¹⁰ If civilization fails, the survivor population could consist mainly of subsistence farmers, hunters and gatherers, and other people who procure their own basic needs. These people are, at present, often considered among the world's poorest, but post-catastrophe they may be among the best off.

Another group that could fare well is those who prepared for the catastrophe. This could include inhabitants of military bunkers and continuity of government facilities, citizen survivalists (“preppers”), and inhabitants of refuges designed for this purpose (Hanson 2008; Jebari 2015; Baum et al. 2015b; Turchin and Green 2017). These people could have ample food and water, medical supplies, and other resources needed to survive without civilization, as well as resources that can be used to maintain and/or rebuild civilization, such as seeds, tools, and information. It could also include much larger populations if there are large-scale catastrophe preparations. At present, most institutional catastrophe preparation is for smaller, more frequent catastrophes such as earthquakes, hurricanes, and sub-pandemic disease outbreaks. However, there are proposals for preparations for large-scale preparation, such as in food stockpiles or

¹⁰ The resilience of civilization to catastrophes is explored in the field of global systemic risk (e.g., Helbing 2013; Baum and Handoh 2014; Centeno et al. 2015).

alternative means of food production to maintain food through catastrophes that disrupt traditional agriculture (Denkenberger and Pearce 2014; Baum et al. 2015a). If proposals along these lines are implemented, survivor populations could be much larger.

For survivor populations, the availability of appropriate information could be a significant factor. The contemporary population is increasingly urban and lacking in the knowledge of how to survive without civilization. However, contemporary civilization has a lot of information documented in libraries, bookstores, and online. While the internet may not be available post-catastrophe, many libraries and bookstores contain basic information about such matters as how to purify water or grow food. Some libraries and bookstores will even have books that are specifically on the topic of how to survive catastrophes (e.g., Dartnell 2014; Denkenberger and Pearce 2014). If survivors can access this information, their prospects could significantly improve.

The availability of food, water, information, and other resources can depend heavily on the specifics of the catastrophe. For example, some catastrophes could cause major disruptions to agriculture (Section 4.3). Nuclear wars are likely to destroy urban areas in targeted countries. Pandemics will leave built infrastructure intact, and may have limited effect in some areas, such as remote islands and inland areas inhabited by uncontacted peoples.

4.2 Successive Generations

If a population is able to survive the immediate catastrophe aftermath and the subsequent years of potentially turbulent conditions, then its subsequent trajectory will depend on, among other things, its ability to produce additional generations of people. If the initial survivor population is too small, or insufficiently healthy, it may fail to remain viable. The population would also need a suitable age and gender distribution. Note that what matters here is not the total worldwide survivor population, but the survivor population(s) that are in sufficient proximity to reproduce. Small survivor populations in geographically distant locations, such as scattered small bunkers or survivalist camps, may contribute little to future generations.

In conservation biology and related fields, the concept of minimum viable population refers to the smallest isolated population able to survive through genetic and environmental changes, with high (generally greater than 90%) probability, for many generations into the future (Shaffer 1981:132). The probability of surviving and thriving increases with the size of the population. Estimates between 150 and 40,000 have been proposed for the minimum viable human population (Lynch et al. 1995; Impey 2015), though some studies argue against any fixed minimum number (Flather et al. 2011; 2016). Some historical analysis of human genetics suggest that there was a genetic bottleneck of as few as 1,000 to 3,000 individuals (Harpending et al. 1998; Li and Durbin 2011; but see Sjödin et al. 2012). Isolated bands of perhaps just 70 humans may have originally colonized both the Americas and Polynesia (Murray-McIntosh et al. 1998; Hey 2005), suggesting a minimum viable population of roughly 100 to 500 given a favorable environment (Hanson 2008; Daily et al. 1993) such as a dedicated refuge for surviving global catastrophes (Hanson 2008; Baum et al. 2015b; Turchin and Green 2017).¹¹

¹¹ It may be possible to recreate a viable human population with advanced synthetic biology technology following a human extinction event (Yampolskiy 2016).

However, in an unstructured and potentially inhospitable post-catastrophe environment, such a population may have a low probability of long-term survival.

This research suggests that larger survivor populations are more likely to succeed, especially in unfavorable post-catastrophe conditions. A population that is just large enough for genetic viability may face other ongoing threats, such as disease, natural disasters, or internal strife. The corresponding trajectory could be low for a period of time before going to zero. That period of time could involve numerous generations, just as small populations on small, remote islands have persisted in isolation for numerous generations. (Certain catastrophe scenarios, such as extreme pandemics, may leave survivors only on small, remote islands.) However, it is unlikely that such populations could avoid a terminal catastrophe for thousands or millions of years into the future, in contrast with a larger population. Indeed, a small survivor population could die out much sooner, such as due to natural disasters, resource overexploitation, internal strife, and/or other factors.

Alternatively, an initially small population could grow much larger. This is especially likely if the population is on a continent, a large island, or a small island that is sufficiently near a continent or large island that the population can reach the larger landmass. Since this is where the overwhelming majority of the current human population resides, there is a high likelihood of survivors in these places, except perhaps in certain catastrophe scenarios that specifically affect the core population, such as extreme pandemics. If there are small survivor populations on or near continents or large islands, they will have an abundance of land to spread out on, making it feasible to grow their population to much larger numbers. However, whether the population, economy, and other attributes can recover to pre-catastrophe levels depends on other factors, in particular regarding agriculture and industry.

4.3 Agriculture

Long-term trajectories could depend heavily on whether agriculture is maintained through the catastrophe or whether it is redeveloped afterwards. Agriculture has played a central role in human civilization to date, and it is difficult to imagine post-catastrophe populations recovering advanced civilization without agriculture. Thus, analysis of long-term catastrophe trajectories should consider prospects for agriculture.

A major catastrophe could significantly alter how humans obtain food. Modern industrial agriculture requires complex resource inputs, supply chains, and labor pools that could be vulnerable to a variety of large catastrophes. For example, Huff et al. (2015) find that major food shortages could occur from pandemics that disrupt the labor supply. More dramatic disruptions to agriculture could come from catastrophes that send particulate matter into the atmosphere, including nuclear war, volcano eruption, and asteroid collision, which send dust into the atmosphere, lowering surface temperatures and reducing sunlight and precipitation over a period of years (Xia et al. 2015).

Catastrophes that disrupt industrial agriculture could see survivors obtaining food via other means, including temporary food stockpiles, food grown from biomass or fossil fuels, or hunting and gathering (Denkenberger and Pearce 2014; Baum et al. 2015a). Hunting and gathering could be especially viable if the human population crashes, leaving an abundance of naturally growing food readily available. Indeed, there is some evidence that at least some hunter-gathers are able to meet their basic needs with

relatively few hours of effort, though this is a point of debate in anthropology (e.g., Bird-David 1992). Furthermore, hunter-gathers may be especially likely to survive some catastrophes, since they often are among the remote uncontacted peoples that would likely survive catastrophes such as pandemics. Thus, despite the dominance of agriculture among pre-catastrophe populations, survivor populations could be hunters and gatherers.

For catastrophes that leave a larger portion of the population intact, agriculture is more likely to endure. Much of the large contemporary urban population has little firsthand experience with either agriculture or hunting/gathering, but it is accustomed to thinking of agriculture as a primary means of food production. Furthermore, information about agriculture is widely documented in libraries and bookstores, as are agriculturally productive species, both of which could be sought out by survivors. This cultural and material legacy makes the survival or rapid redevelopment of agriculture more likely.

Figure 4 presents two illustrative probability distributions of the redevelopment of agriculture. Each version contains three parts: boxes representing the probabilities that agriculture is never lost and that it is never redeveloped (x and z), and a curve showing a probability distribution from the time that the catastrophe occurs to the time the survivor population dies (y_a and y_b). Each version shows it being most likely that agriculture is lost and later recovered and least likely that the survivor population dies without recovering agriculture (i.e., $z < x < y_a/y_b$). Figure 4a shows the survivor population being most likely to recover agriculture shortly after the catastrophe and successively less likely to recover agriculture as time goes by, whereas Figure 4b shows the opposite. Figure 4 is presented for illustration only and should not be treated as a best-guess probability distribution estimate. For example, a best-guess would probably not show the y_b curve as the mirror image of the y_a curve: instead of having the same curvature, it would likely have a more gradual curvature.

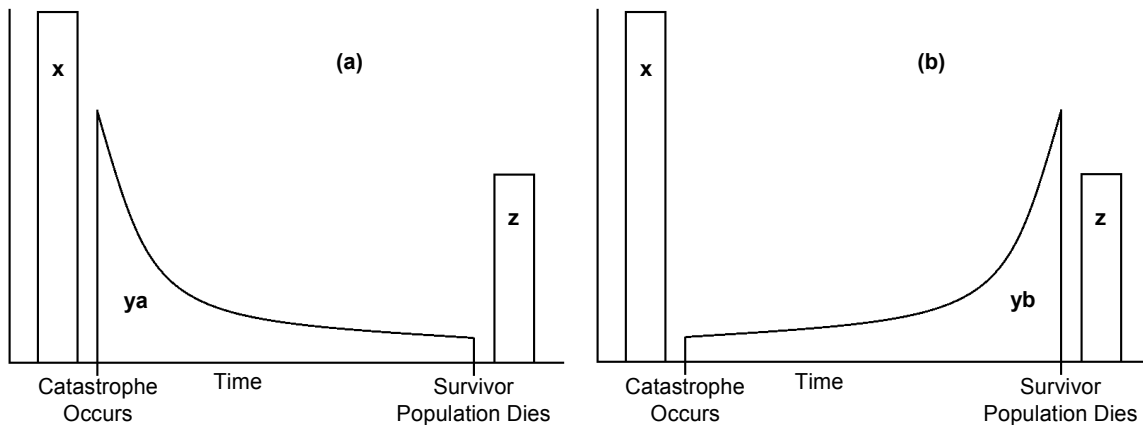


Figure 4. Illustrative probability distributions of post-catastrophe agriculture. The area x is the probability agriculture is never lost; y_a and y_b are the probability it is lost and redeveloped; and z is the probability it is lost and never redeveloped. The curves at y_a and y_b show the probability per unit time of agriculture being redeveloped.

A variety of factors can inform the crafting of agriculture redevelopment probability distributions along the lines of Figure 4. Some of these point to it being more likely that agriculture would be redeveloped sooner, as in Figure 4a. The first post-catastrophe generations would benefit from strong cultural memories of agriculture, including how it

works and the fact that it can produce lots of food. This could lead them to seek to redevelop agriculture and could also facilitate their success. Furthermore, they may have access to important crop and livestock species. These species were developed over thousands of years to maximize productivity and nutritional value, and in this regard are much better than the ancient varieties. Some more recent varieties, such as RoundUp Ready® crops designed to resist a chemical herbicide, may fail to produce significant yields in the absence of modern agriculture, but heirloom varieties would still be considerably more productive than what was available to early humans. These species would likely be more available immediately following the catastrophe; as time goes by, absent human intervention, they may die out or be lost. The availability of grains may be particularly important. Because grains are suitable for long-term storage and transportation, they are easily taxed and thus support the formation of states. Fortunately, modern grains are widely available across all major world regions.

On the other hand, immediately following the catastrophe, agricultural land may be damaged from salination, topsoil depletion, and other effects of modern industrial agriculture. Accessible mineral reserves of phosphorus, a crucial fertilizer, will also be largely depleted (Cordell and White 2011). These conditions would gradually improve over time, improving agriculture prospects for later post-catastrophe generations. If these factors dominate, then the probability distribution for redeveloping agriculture may look more like Figure 4b.

If agriculture does not quickly recover, then prospects for agriculture could depend on fluctuations in Earth's climate. Over the past million years, the climate has fluctuated between cold (glacial) and warm (interglacial) periods, with interglacials occurring every ~100,000 years and lasting for ~15,000 years (Archer 2008).¹² Earth is currently in an interglacial, the Holocene, which began ~10,000 years ago; the previous interglacial, the Eemian, spanned from ~130,000 to ~115,000 years ago (Dahl-Jensen et al. 2013). The rise of agriculture—and the civilization it permits—coincide with the Holocene interglacial, i.e. the last 10,000 years. This may not be a coincidence: the favorable Holocene climate may have been a prerequisite for the development of agriculture (Richerson et al. 2001).

The Holocene's warm climate alone may not have been sufficient. Ice core data suggest that the Holocene has a relatively stable climate compared to previous interglacials (e.g., Petit et al. 1999). Meanwhile, archaeological evidence suggests that the human lineage has had similar cognitive capabilities for ~250,000-300,000 years (McBrearty and Brooks 2000; McBrearty 2013). This suggests that agriculture can only plausibly be developed during some interglacials. As a rough starting estimate, the historical data suggest one third of interglacials are suitable, since three interglacials have occurred in the last 300,000 years, but these data are too sparse to produce reliable conclusions.

Early post-catastrophe climates will differ from early Holocene conditions in one important respect: heightened atmospheric greenhouse gas concentrations. Climate models project anthropogenic greenhouse gas emissions to extend the current interglacial for ~30,000 to ~500,000 years (Archer and Ganopolski 2005; Herrero et al. 2014). This does not necessarily make the early redevelopment of agriculture more likely, because

¹² Patterns with frequencies of ~41,000 years and ~23,000 years are also observed; see e.g. Haqq-Misra (2014).

emissions are making Earth’s climate substantially different than in the early Holocene. Indeed, a warmer world could pose substantial risks to survivor populations, including through ecological changes such as desertification, or greater viability of certain diseases. However, post-catastrophe emissions would presumably be minimal, causing climates to gradually cool.¹³ At some point in the future, before the next glacial period, the survivor population may experience conditions similar to the early Holocene—though perhaps not with the same stability.

Taking these various factors into account, we tentatively estimate agriculture probabilities as shown in Figure 5. Figure 5 is shown as a hand-drawn sketch to emphasize our tentativeness. Figure 5 shows a substantial probability of agriculture not being lost (box x) and a much smaller probability of it never being redeveloped (box z). It further shows some probability of agriculture being quickly redeveloped (the far-left peak of the curve above y). The probability per unit time of the redevelopement of agriculture then quickly falls as survivors struggle under heightened atmospheric greenhouse gas concentrations. The probability per unit time rises again as climates return to something similar to the early Holocene. Finally, the probability per unit time begins an oscillatory pattern corresponding to glacial-interglacial cycles. Figure 5 shows this pattern extending into the indefinite future; in practice, the pattern would eventually fade as Earth shifts into new climate regimes, as caused by the warming and expanding Sun (on time scales of hundreds of millions of years), if not sooner.

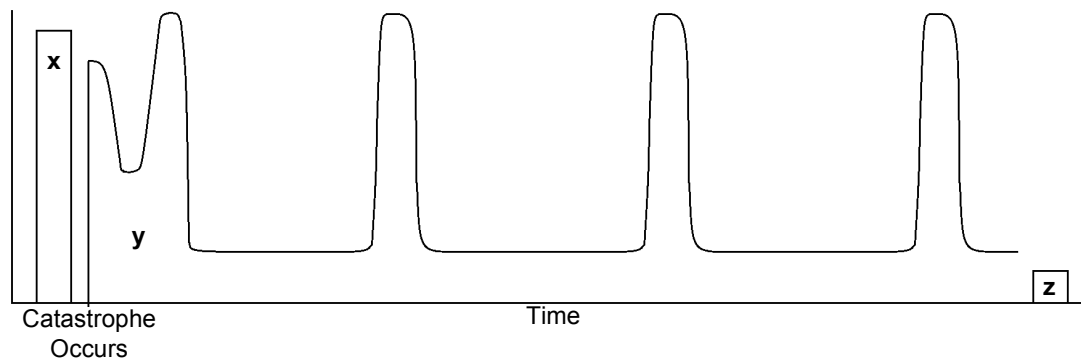


Figure 5. Tentative probability distribution of post-catastrophe agriculture. The area x is the probability agriculture is never lost; y is the probability it is lost and redeveloped; and z is the probability it is lost and never redeveloped. The curve at y shows the probability per unit time of agriculture being redeveloped.

Figure 5 is hardly the only way to plot probabilities of rediscovering agriculture. Indeed, the probabilities depend on several factors, including the catastrophe scenario (including the severity of the catastrophe and the geographic distribution of survivors), pre-disaster preparations, and future greenhouse gas emissions. Likewise, current (pre-catastrophe) civilization can improve prospects in several ways. One is to avoid

¹³ Under most scenarios, there will be some post-catastrophe emissions. For example, carbon dioxide and methane are being released from thawing permafrost (Schuur et al. 2015); this process will continue due to thermal inertia in the climate system unless the catastrophe causes significant ongoing cooling. However, this effect is small relative to the effects from anthropogenic emissions. There may be some scenarios in which the effect is larger, such as scenarios that release the large reservoirs of undersea methane clathrates (on which, see e.g. Buffett and Archer 2004), though these scenarios are less typical.

catastrophes that would be especially disruptive to agriculture, thereby increasing x . Another is to prepare for the redevelopment of agriculture, such as by preserving heirloom agricultural species, thereby increasing y , especially in early post-catastrophe years. Current civilization can also affect the prospects for redeveloping agriculture by increasing or reducing greenhouse gas emissions. However, the net effect of emissions on the probability of redeveloping agriculture is unclear.

4.4 From Agriculture to Industry

In addition to agriculture, modern civilization also depends heavily on industry, as do prospects for technological transformation and space colonization. The current human civilization developed industry after agriculture, and for good reason: agriculture enables the food surplus needed to develop and work in industry. Likewise, a catastrophe that loses agriculture would likely also lose industry, and industry would only be redeveloped after agriculture, whereas a catastrophe that loses industry would not necessarily also lose agriculture. Long-term catastrophe trajectories will thus depend on whether industry is maintained through the catastrophe or whether it is redeveloped afterwards, for which probabilities could be drawn analogously to those in Figure 4.

The short time—just a few thousand years—between the development of agriculture and the development of industry may suggest that if agriculture redevelops, industry is likely to soon follow.

However, there are reasons to be skeptical of this. First, there is only one data point available: industry developed only once, in Britain, and then spread around the world.¹⁴ One data point does not make for reliable analysis: perhaps typical periods from agriculture to industry are much larger or much smaller than a few thousand years. Second, the historical path from agriculture to industry was circuitous. The Chinese, Greeks, Indians, and Romans were all intellectually advanced agricultural civilizations with large trading networks, but they never developed industry. This suggests that the development of industry could have been less of an inevitability and more of a historical coincidence.

Histories of the industrial revolution offer several explanations. Allen (2009) emphasizes the importance of Britain having expensive labor yet inexpensive energy and financial capital, which combined to stimulate the invention and spread of industrial technologies. Mokyr (2009) emphasizes the Enlightenment and scientific revolution, the fruits of which fed entrepreneurs who were empowered by Britain's institutional structure. Crafts (2010) proposes that both sets of factors may have been important. If the previous industrial revolution was indeed contingent on a long list of factors, prospects for the post-catastrophe redevelopment of industry may be poor, because it is relatively unlikely that a multitude of factors would be present concurrently within a survivor population.

Survivor populations would face substantially different resource availabilities. Some of the historical 'low-hanging fruit' will no longer be available. For example, the cheapest and most accessible fossil fuels have already been burned; if these were crucial for Britain's industrial revolution, then post-catastrophe industry may be technologically or economically infeasible. Similarly, the cheapest and most accessible deposits of metal

¹⁴ The fact that industry only developed once may not mean much if it spreads so quickly that multiple independent developments of industry are unlikely.

ores have also been extracted. These factors all suggest a lower probability for the recovery of industry.

Other factors point in the opposite direction. For energy, wind and hydro will be available and are relatively easy to tap. Hydro power in particular has a high energy return on investment, at least under modern conditions (Hall et al. 2014), and has a large global energy production potential (Zhou et al. 2015), making it a promising energy source for recovering industry. Additionally, while major metal deposits will be depleted, these metals themselves will still be present. Indeed, much of the metal has been processed into more usable forms (e.g., steel instead of iron ore) and is concentrated in cities and landfills, the ruins of which could be a good site for post-catastrophe mining. Some of these materials may not be readily usable, such as damaged plastics or certain metals. For instance, even medieval European forges, which lacked bellows systems, were unable to reach temperatures sufficient to melt iron. Nonetheless, some materials are likely to be usable. And if nothing else, the ruins of industry could provide an enduring proof-of-concept and inspiration to redevelop industry.

Given the different conditions faced by survivor populations, it is likely that any civilization that reemerges post-catastrophe would be different from the current one. If resource depletion is the dominant factor, then successive civilizations would tend to be smaller. Alternatively, if the persistent accumulation of knowledge, concentration of metals and other resources in cities, and optimization of agricultural species is the dominant factor, then successive civilizations would tend to be larger. Our present expectation is that successive civilizations will tend to be smaller, due mainly to the depletion of fossil fuels, but we caution that conclusions on this matter cannot be made with confidence at this time.

It is more difficult to project post-catastrophe labor, intellectual, and economic conditions. The catastrophe could decimate the labor supply, but it could do the same for labor demand, with an unclear net effect on the cost of labor. Early post-catastrophe populations may retain some Enlightenment and scientific intellectual tendencies, though this is less likely if survivors hail mainly from pre-catastrophe populations of hunter-gatherers and subsistence farmers. Alternatively, survivors could view industry as a cause of the catastrophe and therefore something to be rejected. If industry is not redeveloped within the first few post-catastrophe generations—for example, if it must wait until a future interglacial period—then these social conditions become especially opaque.

Taking these various factors into account, we tentatively estimate industry probabilities as shown in Figure 6. Figure 6 specifically shows the probability distribution of the recovery of industry at different times given the presence of agriculture. It shows a relatively small probability that industry will not be lost (x), corresponding to catastrophe scenarios in which neither agriculture nor industry are lost. It shows comparable probabilities for industry being or not being recovered, given its loss (y and z). The large z (especially compared to z in Figure 5) is due to the possible historical contingency of the British industrial revolution, as well as the possible narrow time window for industry to redevelop (a few thousand years during interglacials). Finally, Figure 6 shows that industry is likely to be redeveloped sooner after the catastrophe occurs (the curvature of y), due to the social and intellectual advantages that may be retained by the survivor population immediately following the catastrophe.

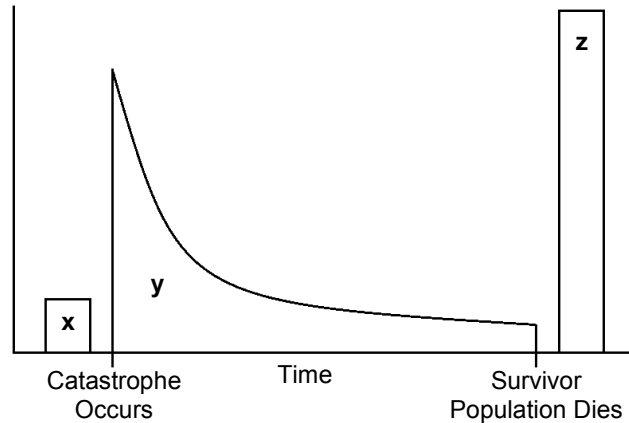


Figure 6. Tentative probability distribution of post-catastrophe industry. The area x is the probability industry is never lost; y is the probability it is lost and redeveloped; and z is the probability it is lost and never redeveloped. The curve at y shows the probability per unit time of industry being redeveloped.

5. Sketches of Technological Transformation and Astronomical Trajectories

A hallmark of technological transformation and astronomical trajectories is that they could both enable rapid expansion of human civilization. (Technological transformation could also lead to catastrophe.) The corresponding trajectories may be similar for both technological transformation and astronomical trajectories and are thus worth considering together before diving into the details of each type of trajectory.

Figure 7 presents several possible trajectories. Each trajectory begins with exponential growth and then goes on a different path. There are two sets of trajectories, one of which lags the other. The leading trajectory shows 2^t during the exponential growth phase, while the lagging trajectory shows 2^{t-n} , with t denoting time and n being an arbitrary finite positive number. The lagging trajectory thus shows delayed technological transformation or space colonization. Figure 7a shows exponential growth continuing indefinitely. Figure 7b shows exponential growth continuing until some fixed size limit is reached, which could be some sort of carrying capacity. Figure 7c shows exponential growth continuing until some fixed point in time is reached. Finally, Figure 7d shows exponential growth continuing until some fixed point in time is reached, at which point civilization crashes. A delay in technological transformation or space colonization results in a long-term reduction in the trajectory in all cases except 7b.

These are not the only possible trajectories. For example, growth could follow a non-exponential curve, or it could taper off gradually, such as in a logistic curve. Trajectories could also oscillate, similar to the rises and falls of civilization in catastrophe and recovery trajectories. Or, trajectories could combine elements from multiple curve types, for example if civilization stops growing as in (b) or (c), and then crashes as in (d), or if civilization branches into different populations with different trajectories. Figure 7 is not intended to be exhaustive, only illustrative of some important possible types of trajectories.

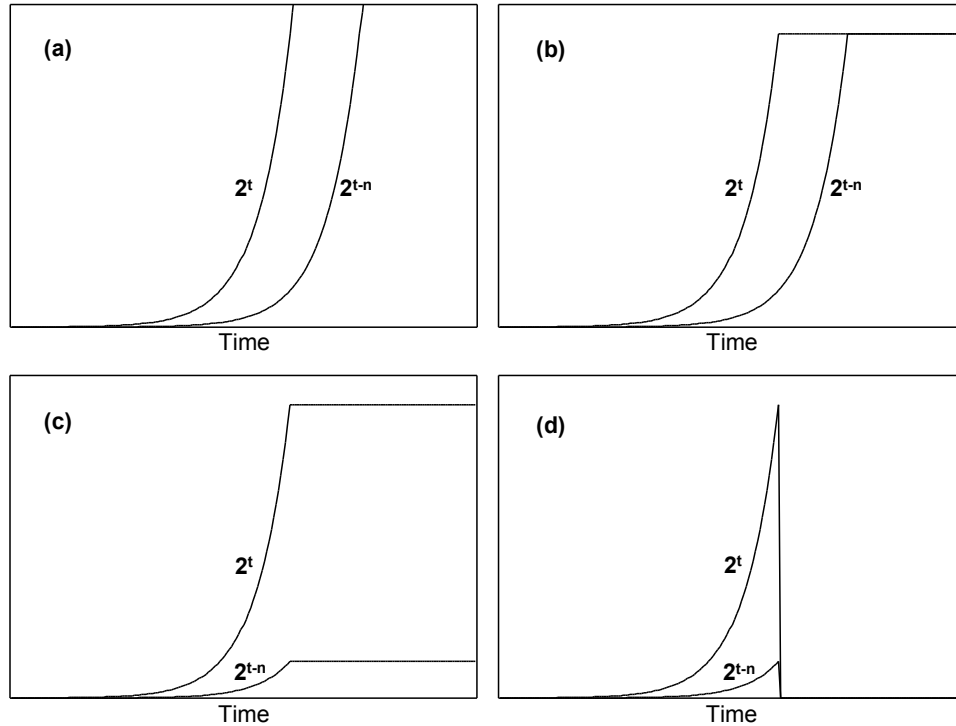


Figure 7. Illustrative technological transformation and astronomical trajectories, including baseline and delayed (by n units of time) exponential growth trajectories for four trajectory types.

6. Technological Transformation Trajectories

Human civilization is currently going through a period of rapid technological change. The status quo trajectory is likely to bring much technological progress. Because of this, it can be difficult to distinguish between status quo and technological transformation trajectories: the status quo trajectory would seem to be one of at least some technological transformation. However, for the technological transformation trajectories, we have in mind an even more profound set of technologies that, if developed, would put human civilization in a fundamentally different place than what may be reached via status quo trajectories. These technologies are often associated with the concept of “technological singularity” because they portend a state change in human civilization (Sandberg 2010). The change can be good or bad, depending on the details of what precisely happens, and depending on one’s ethical perspective. Indeed, compared to status quo and catastrophe trajectories, the evaluation of technological transformation trajectories can be especially sensitive to subtle differences in ethics.

To illustrate technological transformation, this section describes several classes of potential breakthrough technology: nanotechnology, biotechnology, and artificial intelligence. These are not necessarily the only potential breakthrough technologies, but they are among the most widely discussed and perhaps also the most plausible. We review these technologies to convey a sense of the transformation that may be possible and then analyze implications for long-term trajectories. The discussion focuses on trajectories involving technological catastrophe or positive outcomes on Earth; positive outcomes in outer space are discussed in section 7 on astronomical trajectories.

6.1 Nanotechnology

Nanotechnology refers broadly to technology with nanometer-scale components. Nanotechnology is in increasingly wide use with applications in many areas including sunscreens, textiles, and medicine. It is an important area of technology but not particularly transformative. The form of nanotechnology that, if built, is more likely to be transformative, is known as atomically precise manufacturing (APM), which is the assembly of materials with atomic precision. While some doubt the feasibility of APM (Smalley 2001), proof-of-principle can be found in certain biomolecules, such as ribosomes (Freitas and Merkle 2004), and scanning tunneling microscopes, which can perform rudimentary APM (Møller et al. 2017).

If full development of APM is achieved, it could bring a dramatic transformation in human civilization's ability to manufacture a wide range of products. Drexler (2013) terms this "radical abundance", as APM could help eliminate major problems in the fields of food security, treatable diseases, clean energy, and global warming (via clean energy and the manufacture of devices to remove carbon dioxide from the atmosphere¹⁵). APM could also cause problems, such as via enhanced production of weaponry, though this depends on which weapons are built and how they are used. Finally, APM could greatly facilitate space travel, enabling astronomical trajectories, such as by improving the cost and performance of spaceship materials (Drexler 2013).

6.2 Biotechnology

Biotechnology is another sector that is already large yet still only scratching the surface of its transformative potential. Biotechnology is used heavily for livestock and crop plants, a practice that dates to the early days of agriculture or even earlier, and continues to this day with increasing sophistication. Perhaps the biotechnology application with the most transformative potential is in the modification or "enhancement" of human nature. Biotechnology is already in use in such things as cochlear implants, which enable the deaf to hear, and in pharmaceutical cognitive stimulants such as modafinil. Future biotechnologies could turn humans into something much more physically and cognitively capable (Bostrom and Sandberg 2009). In doing so, it could transform human civilization into a civilization of one or more new types of beings that can substantially outperform humans in many ways.

6.3 Artificial Intelligence

At the time of this writing, AI is going through a renaissance, with machine learning algorithms powering advances in many sectors including transportation, medicine, finance, military (Maas et al. 2017), and much more. However, current AI is modest in comparison to some expectations of future AI. In particular, it has been proposed that AI could eventually surpass humanity in important respects, with transformative consequences (Miller 2012; Bostrom 2014; Hanson 2016; Sotala 2017). Such AI could be superintelligent, with much-greater-than-human intellectual capacity, and/or superpowerful, with much-greater-than-human capacity to effect change in the world. Depending on the details of its design, such AI could be transformative for human civilization by solving many of its problems, creating massive new opportunities, or,

¹⁵ For more on carbon dioxide removal, see e.g. Lackner et al. (2012); Tavoni and Socolow (2013).

alternatively, by causing mass destruction (Miller 2012; Bostrom 2014; Sotala and Yampolskiy 2015; Yampolskiy 2015).

6.4 Combinations of Technologies

Nanotechnology, biotechnology, and AI could combine in transformative ways (Eth 2017). Nanotechnology could transform the production of biotechnology, enabling mass adoption of transformative biotechnologies. Biotechnology might facilitate the development of detailed ‘brain emulations’ which accurately simulate human beings, creating one form of AI (Hanson 2016). In turn, AI could facilitate the development of nanotechnology and biotechnology. The AI-biotechnology intersection also includes the prospect of mind uploading, in which human minds are ‘uploaded’ into digital computers, enabling a radically different form of human civilization. Other technologies may also play major roles. Our brief discussion of possible technologies is not intended to be a comprehensive overview—it is intended simply as an overview of the sorts of ways in which technology could transform human civilization.

6.5 Technological Catastrophe Trajectories

The aforementioned technologies could result in several types of trajectories. They could cause catastrophe, such as by nanotechnology being used to mass produce major weaponry, biotechnology being used to engineer deadly pathogens, or AI getting out of control and harming people. Catastrophes from transformative technologies could be considerably more severe than their counterparts involving natural processes or regular technologies: nanotechnology could enable larger wars, biotechnology could enable more severe pathogens, and highly capable AIs could produce uncontrollable disasters. Thus, catastrophes from transformative technologies appear more likely to cause extinction or the loss of agriculture and/or industry. However, for catastrophes of a given severity, the trajectories themselves are likely to be broadly similar regardless of whether the catastrophe is caused by transformative technology or something else.

There is one exception: trajectories involving negative value. Some attributes of human civilization cannot be negative, including population and certain measures of economic production.¹⁶ Others, however, potentially can be. An important one is subjective experience or quality of life. For a given person, subjective experience is negative when, all else equal, they would rather be unconscious—not have any experience at all.¹⁷ This occurs, for example, in medical situations in which anesthetics are used. If a person’s aggregate life experience is negative, that person would arguably

¹⁶ An example of a measure of economic production that cannot be negative is the amount of money spent on goods and services—people cannot spend negative amounts of money. This measure is seen, for example, in the International Monetary Fund’s definition of gross domestic product as “the monetary value of final goods and services—that is, those that are bought by the final user—produced in a country in a given period of time” (<http://www.imf.org/external/pubs/ft/fandd/basics/gdp.htm>). Other measures of economic production can be negative, for example when it is measured relative to initial resource inputs, in which case economic production could produce something less valuable than the initial inputs.

¹⁷ More precisely, subjective experience is negative when a person would evaluate that experience as being worse than the absence of experience, with both being evaluated on their own and not in the context of broader circumstances. There are circumstances in which a person may prefer to have a negative experience. For example, a person may prefer to remain conscious during an extreme emergency, such as a battlefield injury, in order to survive and persevere, even though that would mean enduring great pain.

(depending on one's ethical perspective) be better off not living, and euthanasia may be an appropriate option.¹⁸

Fortunately, the current human population would only benefit from anesthesia or euthanasia in uncommon circumstances. The aggregate subjective experience across the total population would appear to be decidedly positive. Furthermore, it is unlikely that this would change under a wide range of status quo trajectories. Even if the status quo changes for the worse, human psychology has considerable adaptive capacity. Thus, studies of subjective wellbeing find that humans are able to maintain a reasonable quality of life even in seemingly bleak circumstances, such as prostitutes in the slums of Calcutta (Biswas-Diener and Diener 2001).¹⁹

Transformative technologies could bring negative aggregate human subjective experience (Sotala and Gloor 2017). This could occur, for example, by empowering someone or something to push human lives in unbearable ways, such as via mass enslavement or sadism. If transformative technologies are controlled by a small portion of the human population, then they could seek their own benefit to the detriment of everyone else. Alternatively, if autonomous technologies themselves gain control, then they could inflict massive harm. For example, some have worried that an AI designed to improve subjective experience could inadvertently do the opposite, due to subtleties of human psychology that are hard to explain to AI (e.g., Muehlhauser and Helm 2012). If the aggregate would be net negative, then the corresponding trajectory would appear as in Figure 7, except flipped vertically, with negative values instead of positive values on the vertical axis.

Alternatively, transformative technologies could massively improve conditions. They could reduce or even eliminate poverty, disease, environmental degradation, and other major challenges, resulting in excellent versions of status quo trajectories. Transformative technologies could also bring advances that go well beyond the status quo. Some of these involve space colonization, which is discussed further below. However, even just on Earth, much is possible. For example, if human minds are uploaded into computers, they could experience what to them feels like thousands of years in just a few years of conventional calendar time (Hanson 2016).²⁰ Alternatively, and perhaps even more controversially, it may be possible for transformative technology to outright replace humanity with something superior, thereby bringing a massive improvement to the aggregate value on Earth. Thus, even without space colonization, transformative technology could bring a much-greater-than-status quo trajectory.

Still, over the long-term, any civilization that is only on Earth will presumably have certain limits. Earth is limited in volume, in the mass and atomic distribution of available matter, in the quantity of incoming solar radiation, and in its inhabitable lifetime.²¹ Thus,

¹⁸ The concept of negative subjective experience is commonly discussed in the context of negative utilitarianism and related ethical frameworks (e.g., Smart 1958; Griffin 1979; Benatar 2006; Baum 2008).

¹⁹ A possible exception is for populations under highly repressive regimes, especially totalitarian and genocidal regimes, which may create negative subjective wellbeing at larger scales. Additionally, as discussed in Section 4, the total current value may be negative if non-human animal welfare is also accounted for.

²⁰ Of course, thousands of experienced years of negative subjective welfare would be a massive loss, not a massive improvement. The possibility of mind uploading heightens the importance of ensuring positive welfare.

²¹ These parameters may differ from those considered for near-term limits to growth without technological transformation. Nonetheless, they appear to provide strict upper limits for what a more advanced

without space colonization, the long-term trajectory could grow to high levels, though not as high as is possible in astronomical trajectories, and it would go to zero when Earth becomes uninhabitable. Furthermore, it would go to zero at a similar time as status quo trajectories, if both trajectories make it that long. While a civilization with technological transformation may survive somewhat longer, presumably it would die out as the Sun becomes so large that it engulfs the planet. Thus, without space colonization, technological trajectories would likely look something like Figure 7b, except with an eventual decline to zero, as in Figure 7d. Therefore, we should expect the largest and most long-lasting civilizations to come from space colonization.

7. Astronomical Trajectories

The status quo human civilization has already done some basic space exploration, and there is active discussion of establishing permanent colonies on extraterrestrial bodies, especially on Mars.²² However, these space missions are dependent on human civilization on Earth, and likewise are quite small relative to the rest of human civilization. If missions like these will continue to characterize human activity in outer space, then they will have minimal effect on the long-term trajectory of human civilization. In particular, the missions will not significantly increase the total size of human civilization, nor will they increase civilization's longevity. As long as they are dependent on Earth, when Earth civilization eventually dies out, so too will they.

Self-sufficient space colonies could persist for longer. The simplest of these would be self-sufficient versions of the current proposals for small colonies. Perhaps the colonies will be dependent on Earth at first, but gradually become self-sufficient. In that case, they may be able to survive beyond Earth's inhabitable lifetime if they are located further than Earth from the Sun, such as on Mars or certain large moons such as Ganymede or Callisto, or around other stars. If there are only a few such colonies, and if they are of the small size of the colonies currently being proposed, then they would amount to a small tail at the end of the overall long-term trajectory.

The astronomical opportunity greatly increases if colonies can expand themselves on their respective extraterrestrial bodies, just as human civilization has expanded on Earth.²³ This could come, for example, via terraforming, in which habitable atmospheres are engineered on extraterrestrial bodies, and then expanding from small, primarily indoor colonies to outdoor habitation across the entire celestial body. Alternatively, if human minds have been uploaded into digital computers or replaced by AI, then terraforming may be unnecessary for forming large extraterrestrial colonies. In either case, such colonies could come to rival Earth in size, multiplying the size of human civilization by (approximately) the number of colonies.

If extrasolar bodies can be reached, the number of colonies could be quite large. Recent exoplanets research estimates 22% of Sun-like stars have Earth-like planets (Petigura et al. 2013), which would mean that the galaxy contains billions to tens of billions of planets that humans could colonize. If human civilization is in digital instead

technological civilization can do on (or with) Earth.

²² An example of such discussion is the Mars One project, though this project has also been criticized (Do et al. 2016).

²³ This sort of advanced spacefaring capacity could also increase the resilience of human civilization—for example, by enabling engineering at astronomical scales, which can increase resilience to cosmic explosions such as supernovae and gamma ray bursts (Ćirković and Vukotić 2016).

of biological form, then the opportunity increases even further. Even if only a small fraction of these planets are reached, human civilization could grow much larger.²⁴

This raises the question of how civilization would reach so many destinations. The simplest approach is for humanity to send missions from Earth to each destination. However, with so many destinations, this could strain Earth's resources. A more viable approach would be for colonies to self-replicate, with early colonies launching additional colonies as human civilization gradually spreads out across the cosmos. Indeed, using self-replicating tools could enable space colonization at even intergalactic scales (Armstrong and Sandberg 2013).

The resulting growth pattern may approximate an exponential trajectory, with each colony producing new colonies at some rate. Indeed, prior studies of the hypothetical growth of civilizations in the galaxy have often assumed exponential growth. This assumption is central to the Fermi paradox: if extraterrestrial civilizations expanded exponentially, humans are likely to have observed them. Perhaps there are no extraterrestrial civilizations, in which case it may be feasible for human civilization to expand exponentially. Or perhaps they could not sustain exponential growth at galactic scales, in which case human civilization may also not expand so rapidly across the cosmos (Haqq-Misra and Baum 2009).²⁵

There is a geometric reason to expect astronomical expansion at a less-than-exponential trajectory. First, assume that civilization has already maximized value per unit volume, which seems plausible for a civilization that is sufficiently advanced to be able to spread across the cosmos. Second, assume that the civilization is expanding at the maximum possible rate, which may be the speed of light or some other rate given its capacity for expansion. Given that the civilization starts from a small point in space (Earth in the case of human civilization, Earth being a small point in space relative to the rest of the cosmos), the size of civilization would be at most a sphere whose radius increases at the above-mentioned maximum possible rate. With fixed value per volume, this implies that the value of civilization is increasing at a cubic rate, noting that the volume of a sphere is $(4/3)\pi r^3$.

But suppose that human civilization manages to grow exponentially across the cosmos. What happens next? Some of the possibilities are sketched in Figure 7. (Analogous sketches can be made for growth at cubic or other rates.) In astronomical terms, 7a would be if expansion across the cosmos never slows; 7b would be if a carrying capacity is reached, such as if civilization reaches all the habitable planets in the galaxy or all the habitable galaxies in the universe and can expand no further; 7c would be if there is a time when expansion cannot continue, such as if the expansion of the universe ensures some galaxies are pushed too far apart for civilization to reach any more of them; and 7d would be if space colonies fail at some time, such as if the cosmos becomes uninhabitable due to proton decay, heat death, or some other long-term physical process.

Current understanding of the cosmos suggests that 7d, or a variant of 7d, is most likely: it is unlikely that civilization can persist in the cosmos indefinitely. For example, it is expected that, in about 10^{14} years, all stars will stop shining, and in about 10^{36} years,

²⁴ Still further opportunity could come from constructing "Dyson swarms" around stars to collect most of their radiation (e.g., Armstrong and Sandberg 2013).

²⁵ Although for a counter-argument, which suggests that a colonization project for the reachable universe would be a relatively simple task for a star-spanning civilization of a given level, see Armstrong and Sandberg (2013).

proton decay will end matter in its current form (Adams 2008). This is a large but ultimately finite duration. It has been proposed that civilization may be able to move to other universes (Kaku 2005), in which case indefinite survival may be possible, but this physics is speculative and appears unlikely, given what is currently known.

Supposing that human civilization will eventually end, this leaves the question of what may happen in the interim. It is likely that interim trajectories will look broadly similar to 7b for at least an interior geographic core of the civilization. This is because any given finite region of the cosmos presumably has some finite carrying capacity for human civilization, and that an exponentially expanding civilization is likely to reach that carrying capacity long before the region becomes uninhabitable, prompting a crash as in 7d.

After a region's carrying capacity is reached, the trajectory will not necessarily stabilize. The trajectory could decline due to resource depletion and/or competition between populations. Competition on astronomical scales could be analogous to competition between nations or other geographic regions on Earth, potentially pitting planets or star systems against each other; the latter could result in actual star wars.²⁶ Whereas in the initial growth of civilization, newly colonized regions would lack competition and thus could be developed for the sole benefit of the colonizing population, in later times, the emergence of competition could cause populations to divert resources to defense and conquest, thereby reducing the total size of civilization. The magnitude of the reduction could depend on factors such as the viability of conquest at astronomical scales and the capacity of astronomical scale civilization to endure attack.

At the geographic outskirts of civilization, the trajectory could look more like 7a or 7b. 7a would persist if colonization is slow relative to the size of the accessible cosmos, such that human civilization never fills up the entire region that it potentially could. In this case, the total long-term trajectory would look like 7d.

Finally, it should be noted that space colonization could be net negative according to some metrics (Sotala and Gloor 2017). Space colonization could expand net negatives on Earth achieved via technological transformation (Section 6.5). Alternatively, space colonization could bring net negatives even when conditions on Earth are net positive. Outer space is a less hospitable environment than Earth, which could make net negatives more likely. It is further conceivable that the outer space environment may be conducive to oppressive regimes and/or violent conflict that also yield net negatives. As with technological transformation, if space colonization would bring net negatives, then the trajectories would look like those in Figure 7, except flipped vertically.

8. Discussion and Conclusion

This paper has presented a range of ethical and empirical aspects of the long-term trajectories of human civilization. This is in contrast with most prior research, which focuses on near-term trajectories, perhaps due to a mix of ethical favoritism for the near-term and scientific reluctance to take on the seemingly more speculative nature of long-term trajectories. In this paper, we hope to have shown that the long-term trajectory is both an important topic and one on which intelligent things can be said. We also hope to

²⁶ The astronomical distances involved here make this situation not exactly analogous to competition on Earth. For example, communication is much slower at distances of light years or more, which precludes certain types of military tactics.

have produced some basic findings about long-term trajectories and demonstrated the sorts of research that can advance the topic.

One basic finding is that the status quo human civilization appears unlikely to persist over the long-term. Instead, it is likely to either end catastrophically or expand dramatically. Over the very long term, it appears likely that human civilization will cease to exist. However, it may be a very long time until that happens, especially if civilization colonizes space and of course only if it avoids the near-term threats to human extinction. In the meantime, civilization could follow a wide range of trajectories, from mere survival without agriculture or industry to extensive colonization of the cosmos.

Which of these trajectories will occur can be important for contemporary decision making. For example, recall the five potential claims about contemporary decision making introduced in the Section 2 discussion of time-neutral total utilitarianism:

- (A) Current decisions should prioritize reducing the risk of human extinction, because extinction would result in the loss of all future generations.
- (B) Current decisions should prioritize reducing the risk of major catastrophes that would lead to the permanent loss of advanced human civilization.
- (C) Current decisions should prioritize expediting technological breakthroughs and ensuring that they would improve welfare.
- (D) Current decisions should prioritize expediting space colonization and ensuring that it would improve welfare.
- (E) Current decisions should prioritize improving near-term welfare, because long-term trajectories will be the same either way.

Given a time-neutral total utilitarianism ethical framework, these claims can be evaluated in terms of the effect that contemporary decisions have on long-term trajectories. In light of this paper's analysis, the following can be said about the sorts of circumstances in which each claim would hold:

- (A) There are risks to human extinction that contemporary decisions can affect, and risks of sub-extinction catastrophes would not affect the long-term trajectory of human civilization. Sub-extinction catastrophes would have no long-term effect if they would cause no delay in space colonization, which is unlikely, or if delays in space colonization would have no effect on long-term trajectories.
- (B) There are risks to the loss of advanced human civilization that contemporary decisions can affect, and survivors of such catastrophes are unlikely to recover advanced civilization.
- (C) There are opportunities to influence technological breakthroughs.
- (D) There are opportunities to influence space colonization.
- (E) There are no significant opportunities to affect major catastrophes, technological breakthroughs, or space colonization.

The finding that status quo trajectories are unlikely to persist could suggest that claim (E) is false. If human civilization will end up in some non-status quo trajectory, then perhaps there are opportunities to affect this trajectory. This is not necessarily the case: human civilization may end up in a certain non-status quo trajectory regardless of what

actions are currently taken. For example, perhaps a looming catastrophe will destroy civilization, and there is nothing that anyone can do about it. However, we do not believe this to be the case. We see a wide range of opportunities in particular to affect some combination of catastrophic risks, potentially transformative technologies, and space colonization. Which opportunities are best to pursue will depend on both ethical factors and the specifics of the opportunities available to different people.

Within the context of catastrophic risks, we can compare claims (A) and (B). A number of catastrophic risks may not result in immediate extinction. This paper's analysis suggests attention to several factors, in particular the post-catastrophe retention or recovery of agriculture and industry, as well as the long-term significance of delays in space colonization. Though tentative, the analysis finds that the success of post-catastrophe human civilization is not guaranteed, with the recovery of industry being a particularly strong impediment. Catastrophe recovery would likely induce delays in space colonization, which could cause long-term trajectory declines. These findings suggest that claim (A) may be false and claim (B) may be true. In other words, contrary to some claims in the catastrophic risk literature, extinction risks may not be categorically more important than large sub-extinction risks. Whether a particular opportunity to reduce an extinction risk is better than a particular opportunity to reduce a large sub-extinction risk will depend on the particulars of the opportunities as well as the relative long-term trajectories.

Likewise, the relative importance of opportunities to reduce catastrophic risks, develop transformative technologies, and launch space missions also depends on a number of particulars. One important variable is timing. For example, if technological breakthroughs or space colonization will occur relatively quickly, this can diminish the importance of many potential catastrophe scenarios, especially where the technologies or space colonies can mitigate the catastrophic risk. Alternatively, an early catastrophe could preclude the development of transformative technology or the launch of space colonies. Additionally, transformative technology may greatly facilitate space colonization, and may be worth pursuing prior to extensive efforts to colonize space. Some of these events could play out on relatively near-term time scales, making them especially important to track and pursue opportunities to influence.

The above discussion pertains specifically to time-neutral total utilitarianism. We emphasize that our discussion of this ethical view is intended as an example, not as an endorsement. It is not our aim in this paper to argue for any particular ethical theory—it is only to show how long-term trajectories can be important according to some ethical theories, and likewise how the study of long-term trajectories should include ethical theory in addition to empirical detail. For the many ethical theories that do value future events, the study of long-term trajectories is highly important. The stakes are extremely large, and there may be a lot that people today can do to have a positive impact. For these reasons, the long-term trajectories of human civilization should be an important focus of both academic scholarship and societal decision-making.

Acknowledgments

This paper is based on a discussion led by Baum on 9 September 2017 at the Workshop on Existential Risk to Humanity, which was hosted by Häggström and Sandberg at Chalmers University of Technology. We thank Catherine Rhodes, Aron Vallinder, and

Johan Wästlund for comments at the discussion. We also thank Jacob Haqq-Misra and two anonymous reviewers for helpful feedback on an earlier version of this paper and Melissa Thomas-Baum for assistance with graphics.

References

- Adams, F. C. (2008), “Long-term astrophysical processes”, in Bostrom, N. and Ćirković, M. M. (Eds.), *Global Catastrophic Risks*, Oxford University Press, Oxford, pp. 33-47.
- Adler, M. (2012). *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. Oxford University Press.
- Agar, N. (2000), *Life’s Intrinsic Value: Science, Ethics and Nature*, Columbia University Press, New York.
- Allen, R. C. (2009), *The British Industrial Revolution in Global Perspective*. Cambridge University Press, Cambridge, UK.
- Archer, D. (2008), *The Long Thaw: How Humans are Changing the Next 100,000 Years of Earth’s Climate*, Princeton University Press, Princeton.
- Archer, D. and Ganopolski, A. (2005), “A movable trigger: Fossil fuel CO₂ and the onset of the next glaciation”, *Geochemistry, Geophysics, Geosystems*, 6(5).
- Armstrong, S. and Sandberg, A. (2013), “Eternity in six hours: Intergalactic spreading of intelligent life and sharpening the Fermi paradox”, *Acta Astronautica*, Vol. 89, pp. 1-13.
- Arrhenius, G. (2000), “An impossibility theorem for welfarist axiologies”, *Economics & Philosophy*, Vol. 16, No. 2, pp. 247-266.
- Arrhenius, G. (2003), “The person-affecting restriction, comparativism, and the moral status of potential people”, *Ethical Perspectives*, Vol. 10, Nos. 3-4, pp. 185-195.
- Asimov, I. (1981), *A Choice of Catastrophes: The Disasters That Threaten Our World*. Ballantine, New York.
- Baum, S. D. (2008), “Better to exist: A reply to Benatar”, *Journal of Medical Ethics*, Vol. 34, No. 12, pp. 875-876.
- Baum, S. D., Denkenberger, D. C., Pearce, J. M., Robock, A. and Winkler, R. (2015a), “Resilience to global food supply catastrophes”, *Environment, Systems, and Decisions*, Vol. 35, No. 2, pp. 301-313.
- Baum, S. D., Denkenberger, D. C. and Haqq-Misra, J. (2015b), “Isolated refuges for surviving global catastrophes”, *Futures*, Vol. 72, pp. 45-56.
- Baum, S. D. and Handoh, I. C. (2014), “Integrating the planetary boundaries and global catastrophic risk paradigms”, *Ecological Economics*, Vol. 107, pp. 13-21.
- Beckstead, N. (2013), *On the Overwhelming Importance of Shaping the Far Future*, Doctoral Dissertation, Department of Philosophy, Rutgers University.
- Benatar, D. (2006), *Better Never to Have Been: The Harm of Coming Into Existence*. Clarendon Press, Oxford.
- Bhat, Z.F. and Fayaz, H. (2011), “Prospectus of cultured meat—advancing meat alternatives”, *Journal of Food Science and Technology*, Vol. 48, No. 2, pp. 125-140.
- Bird-David, N. (1992). “Beyond ‘the original affluent society’: A culturalist reformulation”, *Current Anthropology*, Vol. 33, No. 1, pp. 25-47.
- Biswas-Diener, R. and Diener, E. (2001), “Making the best of a bad situation: Satisfaction in the slums of Calcutta”, *Social Indicators Research*, Vol. 55, No. 3, pp. 329-352.

- Bostrom, N. (2003), "Are we living in a computer simulation?" *Philosophical Quarterly*, Vol. 53, no. 211, pp. 243-255.
- Bostrom, N. (2013), "Existential risk prevention as a global priority", *Global Policy*, Vol. 4, No. 1, pp. 15-31.
- Bostrom, N. (2014), *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press, Oxford.
- Bostrom, N. and Ćirković, M. (2008), *Global Catastrophic Risks*, Oxford University Press, Oxford.
- Bostrom, N. and Sandberg, A. (2009), "Cognitive enhancement: Methods, ethics, regulatory challenges", *Science and Engineering Ethics*, Vol. 15, No. 3, pp. 311-41.
- Buffett, B. and Archer, D. (2004), "Global inventory of methane clathrate: sensitivity to changes in the deep ocean", *Earth and Planetary Science Letters*, Vol. 227, Nos. 3-4, pp. 185-199.
- Centeno, M. A., Nag, M., Patterson, T. S., Shaver, A. and Windawi, A. J. (2015), "The emergence of global systemic risk", *Annual Review of Sociology*, Vol. 41, pp. 65-85.
- Chalmers, D. J. (2005), "The Matrix as metaphysics", in Grau, C. (Ed.), *Philosophers Explore the Matrix*, Oxford University Press, Oxford, pp. 132-176.
- Ćirković, M. M. (2002), "Cosmological forecast and its practical significance", *Journal of Evolution and Technology*, Vol. 12.
- Ćirković, M. M. and Vukotić, B. (2016), "Long-term prospects: Mitigation of supernova and gamma-ray burst threat to intelligent beings", *Acta Astronautica*, Vol. 129, pp. 438-446.
- Ćirković, M. M., Sandberg, A. and Bostrom, N. (2010), "Anthropic shadow: observation selection effects and human extinction risks", *Risk Analysis*, Vol. 30, No. 10, pp. 1495-1506.
- Cordell, D. and White, S. (2011), "Peak phosphorus: Clarifying the key issues of a vigorous debate about long-term phosphorus security", *Sustainability*, Vol. 3, No. 10, pp. 2027-2049.
- Crafts, N. (2010). "Explaining the first Industrial Revolution: Two views", *European Review of Economic History*, Vol. 15, No. 1, pp. 153-168.
- Dahl-Jensen, D., Albert, M. R., Aldahan, A., Azuma, N., Balslev-Clausen, D., Baumgartner, M., et al. (2013), "Eemian interglacial reconstructed from a Greenland folded ice core", *Nature*, Vol. 493, No. 7433, 489-494.
- Daily, G. C., Ehrlich, A. H. and Ehrlich, P. R. (1993), "Optimum human population size", *Race, Poverty and The Environment*, Vol. 4, No. 2, pp. 9-12.
- Dartnell, L. (2014), *The Knowledge: How to Rebuild Civilization in the Aftermath of a Cataclysm*, Penguin, London.
- DeGrazia, D. (2009), "Moral vegetarianism from a very broad basis", *Journal of Moral Philosophy*, Vol. 6, pp. 143-165.
- Denkenberger, D. and Pearce, J. (2014), "Feeding everyone no matter what: Managing food security after global catastrophe", Academic Press, Waltham.
- Do, S., Owens, A., Ho, K., Schreiner, S. and De Weck, O. (2016), "An independent assessment of the technical feasibility of the Mars One mission plan-Updated analysis", *Acta Astronautica*, Vol. 120, pp. 192-228.
- Drexler, K. E. (2013), *Radical Abundance: How a Revolution in Nanotechnology Will Change Civilization*. PublicAffairs, New York.

- Eden, A. H., Moor, J. H., Soraker, J. H. and Steinhart, E. (Eds.) (2012), *Singularity Hypotheses: A Scientific and Philosophical Assessment*. Springer, Berlin.
- Eth, D. (2017), “The technological landscape affecting artificial general intelligence and the importance of nanoscale neural probes”, *Informatica*, Vol. 41, No. 4, pp. 463-470.
- Field, C. B., Barros, V. R., Mach, K. J., Mastrandrea, M. D., van Aalst, M., Adger, W. N., et al. (2014), “Technical summary”, in Field, C. B., Barros, V. R., Dokken, D. J., Mach, K. J., Mastrandrea, M. D., Bilir, T. E., et al. (Eds.), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, U. K., pp. 35-94.
- Flather, C. H., Hayward, G. D., Beissinger, S. R. and Stephens, P. A. (2011), “Minimum viable populations: Is there a ‘magic number’ for conservation practitioners?”, *Trends in Ecology & Evolution*, Vol. 26, No. 6, pp. 307-316.
- Freitas Jr., R. A. and Merkle, R. C. (2004), *Kinematic Self-Replicating Machines*. Landes Bioscience, Austin.
- Frick, J. (2017), “On the survival of humanity”, *Canadian Journal of Philosophy*, Vol. 47, No. 2-3, pp. 344-367.
- Gloor, L. and Mannino, A. (2016), “The case for suffering-focused ethics”, Foundational Research Institute, <https://foundational-research.org/the-case-for-suffering-focused-ethics>.
- Griffin, J. (1979), “Is unhappiness morally more important than happiness?”, *Philosophical Quarterly*, Vol. 29, No. 114, pp. 47-55.
- Hägström, O. (2016), *Here Be Dragons: Science, Technology and the Future of Humanity*, Oxford University Press, Oxford.
- Hall, C. A., Lambert, J. G. and Balogh, S. B. (2014), “EROI of different fuels and the implications for society”, *Energy Policy*, Vol. 64, pp. 141-152.
- Hanson, R. (2008), “Catastrophe, social collapse, and human extinction”, in Bostrom, N. and Čirković, M. M. (Eds.), *Global Catastrophic Risks*, Oxford University Press, Oxford, pp. 363-377.
- Hanson, R. (2016), *The Age of Em: Work, Love, and Life When Robots Rule the Earth*, Oxford University Press, Oxford.
- Haqq-Misra, J. (2014), “Damping of glacial–interglacial cycles from anthropogenic forcing”, *Journal of Advances in Modeling Earth Systems*, Vol. 6, No. 3, pp. 950-955.
- Haqq-Misra, J. D. and Baum, S. D. (2009), “The sustainability solution to the Fermi paradox”, *Journal of the British Interplanetary Society*, Vol. 62, No. 2, pp. 47-51.
- Harpending, H. C., Batzer, M. A., Gurven, M., Jorde, L. B., Rogers, A. R. and Sherry, S. T. (1998), “Genetic traces of ancient demography”, *Proceedings of the National Academy of Sciences*, Vol. 95, No. 4, pp. 1961-1967.
- Helbing, D. (2013), “Globally networked risks and how to respond”, *Nature*, Vol. 497, No. 7447, pp. 51-59.
- Herrero, C., García-Olivares, A., and Pelegrí, J. L. (2014), “Impact of anthropogenic CO₂ on the next glacial cycle”, *Climatic Change*, Vol. 122, Nos. 1-2, pp. 283-298.
- Hey, J. (2005), “On the number of new world founders: A population genetic portrait of the peopling of the Americas”, *PLOS Biology*, Vol. 3, No. 6, e193.

- Huff, A. G., Beyeler, W. E., Kelley, N. S., and McNitt, J. A. (2015), “How resilient is the United States’ food system to pandemics?”, *Journal of Environmental Studies and Sciences*, Vol. 5, No. 3, pp. 337-347.
- Impey, C. (2015), *Beyond: Our Future in Space*, Norton, New York.
- Jebari, K. (2015), “Existential risks: Exploring a robust risk reduction strategy”, *Science and Engineering Ethics*, Vol. 21, No. 3, pp. 541–554
- Johansson-Stenman, O. (2018), “Animal welfare and social decisions: Is it time to take Bentham seriously?”, *Ecological Economics*, Vol. 145, pp. 90-103.
- Kaku, M. (2005), *Parallel Worlds: The Science of Alternative Universes and Our Future in the Cosmos*. Penguin, London.
- Kurzweil, R. (2005), *The Singularity Is Near: When Humans Transcend Biology*, Viking. New York.
- Lackner, K. S., Brennan, S., Matter, J. M., Park, A. H. A., Wright, A., and Van Der Zwaan, B. (2012), “The urgency of the development of CO2 capture from ambient air”, *Proceedings of the National Academy of Sciences*, Vol. 109, No. 33, pp. 13156-13162.
- Lebow, R. N. (2015), “Counterfactuals and security studies”, *Security Studies*, Vol. 24, No. 3, pp. 403-412.
- Leggett, M. (2006), “An indicative costed plan for the mitigation of global risks”, *Futures*, 38(7), 778-809.
- Li, H. and Durbin, R. (2011), “Inference of human population history from individual whole-genome sequences”, *Nature*, Vol. 475, No. 7357, pp. 493-496.
- Lynch, M., Conery, J. and Burger, R. (1995), “Mutational meltdowns in sexual populations”, *Evolution*, Vol. 49, No. 6, pp. 1067-1080.
- Maas, M., Sweijts, T. and De Spiegeleire, S. (2017), *Artificial Intelligence and the Future of Defense: Strategic Implications for Small- and Medium-Sized Force Providers*, The Hague Centre for Strategic Studies, The Hague, <http://hcss.nl/report/artificial-intelligence-and-future-defense>.
- Maher Jr., T. M. and Baum, S. D. (2013), “Adaptation to and recovery from global catastrophe”, *Sustainability*, Vol. 5, No. 4, pp. 1461-1479.
- Mallon, R. (2005), “The deplorable standard of living faced by farmed animals in America’s meat industry and how to improve conditions by eliminating the corporate farm”, *Michigan State University Journal of Medicine and Law*, Vol. 9, pp. 389-415.
- Matheny, J. G. (2007), “Reducing the risk of human extinction”, *Risk Analysis*, Vol. 27, No. 5, pp. 1335-1344.
- McBrearty, S. (2013), “Advances in the study of the origin of humanness”, *Journal of Anthropological Research*, Vol. 69, No. 1, pp. 7-31.
- McBrearty, S. and Brooks, A. S. (2000), “The revolution that wasn’t: A new interpretation of the origin of modern human behavior”, *Journal of Human Evolution*, Vol. 39, No. 5, pp. 453-563.
- Miller, J. (2012), *Singularity Rising: Surviving and Thriving in a Smarter, Richer, and More Dangerous World*, BenBella, Dallas.
- Mokyr, J. (2009), *The Enlightened Economy: An Economic History of Britain 1700-1850*, Yale University Press, New Haven.

- Møller, M., Jarvis, S. P., Guérinet, L., Sharp, P., Woolley, R., Rahe, P. and Moriarty, P. (2017), “Automated extraction of single H atoms with STM: Tip state dependency”, *Nanotechnology*, Vol. 28, No. 7, 075302.
- Moravec, H. (1998), “When will computer hardware match the human brain”, *Journal of Evolution and Technology*, Vol. 1.
- More, M. and Vita-More, N (Eds.) (2010), *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, Wiley, New York.
- Morito, B. (2003), “Intrinsic value: A modern albatross for the ecological approach”, *Environmental Values*, Vol. 12, pp. 317-336.
- Muehlhauser, L. and Helm, L. (2012), “The singularity and machine ethics”, in Eden, A. H., Moor, J. H., Soraker, J. H. and Steinhart, E. (Eds.), *Singularity Hypotheses: A Scientific and Philosophical Assessment*. Springer, Berlin, pp. 101-126.
- Murray-McIntosh, R. P., Scrimshaw, B. J., Hatfield, P. J. and Penny, D. (1998), “Testing migration patterns and estimating founding population size in Polynesia by using human MtDNA sequences”, *Proceedings of the National Academy of Sciences*, Vol. 95, No. 15, pp. 9047-9052.
- Ng, Y. K. (1989), “What should we do about future generations? Impossibility of Parfit’s theory X”, *Economics & Philosophy*, Vol. 5, No. 2, pp. 235-253.
- Ng, Y. K. (1990), “Welfarism and utilitarianism: A rehabilitation”, *Utilitas*, Vol. 2, No. 2, pp. 171-193.
- Ng, Y-K (1991), “Should we be very cautious or extremely cautious on measures that may involve our destruction?”, *Social Choice and Welfare*, Vol. 8, No. 1, pp. 79-88.
- Norton, B. (1984), “Environmental ethics and weak anthropocentrism”, *Environmental Ethics*, Vol. 6, No. 3, pp. 131-148.
- Nozick, R. (1974), *Anarchy, State, and Utopia*, Basic Books, New York.
- O’Malley-James, J. T., Cockell, C. S., Greaves, J. S. and Raven, J. A. (2014), “Swansong biospheres II: The final signs of life on terrestrial planets near the end of their habitable lifetimes”, *International Journal of Astrobiology*, Vol. 13, No. 3, pp. 229-243.
- Ord, T., Hillerbrand, R. and Sandberg, A. (2010), “Probing the improbable: Methodological challenges for risks with low probabilities and high stakes”, *Journal of Risk Research*, Vol. 13, pp. 191-205.
- Parfit, D. (1984), *Reasons and Persons*, Oxford University Press, Oxford.
- Petigura, E. A., Howard, A. W. and Marcy, G. W. (2013), “Prevalence of Earth-size planets orbiting Sun-like stars”, *Proceedings of the National Academy of Sciences*, Vol. 110, No. 48, pp. 19273-19278.
- Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J. M., Basile, I., et al. (1999), “Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica”, *Nature*, Vol. 399, No. 6735, pp. 429-436.
- Portney, P. and Weyant, J. (Eds.) (1999), *Discounting and Intergenerational Equity*, Resources For the Future, Washington, D. C.
- Post, M. J. (2012), “Cultured meat from stem cells: Challenges and prospects”, *Meat Science*, Vol. 92, pp. 297-301.
- Rees, M. (2003), *Our Final Century: Will the Human Race Survive the Twenty-first Century?*, William Heinemann, Oxford.

- Richerson, P. J., Boyd, R. and Bettinger, R. L. (2001), “Was agriculture impossible during the Pleistocene but mandatory during the Holocene? A climate change hypothesis”, *American Antiquity*, Vol. 66, No. 3, pp. 387-411.
- Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E., et al. (2009), “A safe operating space for humanity”, *Nature*, Vol. 461, pp. 472-475.
- Sagan, C. (1983), “Nuclear war and climatic catastrophe: Some policy implications”, *Foreign Affairs*, Vol. 62, pp. 257-292.
- Sagan, C. (2006), *The Varieties of Scientific Experience: A Personal View of the Search for God*, Penguin, New York.
- Sandberg, A. (2010), “An overview of models of technological singularity”, in More, M. and Vita-More, N (Eds.), *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, Wiley, New York, pp. 376-394.
- Scheffler, S. (2018), *Why Worry About Future Generations?*, Oxford University Press, Oxford.
- Schuur, E. A. G., McGuire, A. D., Schädel, C., Grosse, G., Harden, J. W., Hayes, D. J., et al. (2015), “Climate change and the permafrost carbon feedback”, *Nature*, Vol. 520, No. 7546, pp. 171-179.
- Schwartz, J. S. J. (2011), “Our moral obligation to support space exploration”, *Environmental Ethics*, Vol. 33 pp. 67-88.
- Schwartz, J. S. J. and Milligan, T. (Eds.) (2016), *The Ethics of Space Exploration*, Springer, Cham, Switzerland.
- Shaffer, M. L. (1981), “Minimum population sizes for species conservation”, *BioScience*, Vol. 31, No. 2, pp. 131-134.
- Sjodin, P., E. Sjöstrand, A., Jakobsson, M. and Blum, M. G. (2012), “Resequencing data provide no evidence for a human bottleneck in Africa during the penultimate glacial period”, *Molecular Biology and Evolution*, Vol. 29, No. 7, pp. 1851-1860.
- Smalley, R. E. (2001), “Of chemistry, love and nanobots”, *Scientific American*, September, pp. 76-77.
- Smart R. (1958), “Negative utilitarianism”, *Mind*, Vol. 67, pp. 542-543.
- Sotala, K. (2017), “How feasible is the rapid development of artificial superintelligence?”, *Physica Scripta*, Vol. 92, No. 11, 113001.
- Sotala, K. and Gloor, L. (2017), “Superintelligence as a cause or cure for risks of astronomical suffering”, *Informatica*, Vol. 41, No. 4, pp. 389-400.
- Sotala, K. and Yampolskiy, R. V. (2015), “Responses to catastrophic AGI risk: A survey”, *Physica Scripta*, Vol. 90, No. 1, 018001.
- Tavoni, M. and Socolow, R. (2013), “Science and policy of negative emission technologies” (special issue), *Climatic Change*, Vol. 118, No. 1, pp. 1-149.
- Tonn, B. E. (2017), “Philosophical, institutional, and decision making frameworks for meeting obligations to future generations”, *Futures*, Vol. 95, pp.44-57.
- Tonn, B. and MacGregor, D. (2009), “A singular chain of events”, *Futures*, Vol. 41, pp. 706-714.
- Turchin, A. and Green, B. P. (2017), “Aquatic refuges for surviving a global catastrophe”, *Futures*, Vol. 89, pp. 26-37.
- Turner, M. S. and Wilczek, F. (1982), “Is our vacuum metastable?”, *Nature*, Vol. 298, pp. 635-636.

- United Nations (2012), “Livestock primary dataset: World total of animals slaughtered for meat in 2012”, <http://faostat.fao.org/site/569/DesktopDefault.aspx?PageID=569>.
- United Nations (2017a), “World population prospects: The 2017 revision, key findings and advance tables”, United Nations Department of Economic and Social Affairs, Population Division, Working Paper No. ESA/P/WP/248.
- United Nations (2017b). “World population prospects: The 2017 Revision, DVD Edition”, United Nations Population Division, Department of Economic and Social Affairs, <https://esa.un.org/unpd/wpp/Download/Standard/Population>.
- United Nations (2017c). “Probabilistic population projections based on the world population prospects: The 2017 revision”, United Nations Population Division, Department of Economic and Social Affairs, <https://esa.un.org/unpd/wpp/Download/Probabilistic/Population>.
- Weinberg, R. (2008), “Identifying and dissolving the non-identity problem”, *Philosophical Studies*, Vol. 137, No. 1, pp. 3-18.
- Weisbach, D. A. and Sunstein, C. R. (2007), “Introduction: Symposium on intergenerational equity and discounting”, *University of Chicago Law Review*, Vol. 74, No. 1, pp. 1-3.
- Wills, C. (2008), “Evolution theory and the future of humanity”, in Bostrom, N. and Ćirković, M. M. (Eds.), *Global Catastrophic Risks*, Oxford University Press, Oxford, pp. 48-72.
- Wolf, E. T. and Toon, O. B. (2015), “The evolution of habitable climates under the brightening Sun”, *Journal of Geophysical Research: Atmospheres*, Vol. 120, No. 12, pp. 5775-5794.
- Yampolskiy, R. (2015), *Artificial Superintelligence: A Futuristic Approach*, CRC Press, Boca Raton.
- Yampolskiy, R.V. (2016), “On the origin of synthetic life: Attribution of output to a particular algorithm”, *Physica Scripta*, Vol. 92, No. 1, 013002.
- Xia, L., Robock, A., Mills, M., Stenke, A. and Helfand, I. (2015), “Decadal reduction of Chinese agriculture after a regional nuclear war”, *Earth's Future*, Vol. 3, pp. 37-48.
- Zhou, Y., Hejazi, M., Smith, S., Edmonds, J., Li, H., Clarke, L., Calvin, K. and Thomson, A. (2015), “A comprehensive view of global potential for hydro-generated electricity”, *Energy & Environmental Science*, Vol. 8, No. 9, pp. 2622-2633.