**Global Catastrophic and Existential Risks**

**Communication Scale**

*Preprint*

Published in *Futures*

[*https://www.sciencedirect.com/science/article/pii/S001632871730112X*](https://www.sciencedirect.com/science/article/pii/S001632871730112X)

*https://doi.org/10.1016/j.futures.2018.01.003*

*Dec 2017*

Alexey Turchin

Science for Life Extension Foundation, Moscow

David Denkenberger

Global Catastrophic Risk Institute;

Alliance to Feed the Earth in Disasters;

Tennessee State University

**Abstract**. Existential risks threaten the future of humanity, but they are difficult to measure. However, to communicate, prioritize and mitigate such risks it is important to estimate their relative significance. Risk probabilities are typically used, but for existential risks they are problematic due to ambiguity, and because quantitative probabilities do not represent some aspects of these risks. Thus, a standardized and easily comprehensible instrument is called for, to communicate dangers from various global catastrophic and existential risks. In this article, inspired by the Torino scale of asteroid danger, we suggest a color coded scale to communicate the magnitude of global catastrophic and existential risks. The scale is based on the probability intervals of risks in the next century if they are available. The risks’ estimations could be adjusted based on their severities and other factors. The scale covers not only existential risks, but smaller size global catastrophic risks. It consists of six color levels, which correspond to previously suggested levels of prevention activity. We estimate artificial intelligence risks as “red”, while “orange” risks include nanotechnology, synthetic biology, full scale nuclear war and a large global agricultural shortfall (caused by regional nuclear war, coincident extreme weather, etc.) The risks of natural pandemic, supervolcanic eruption and global warming are marked as “yellow” and the danger from asteroids is “green”.

**Keywords**: global catastrophic risks; existential risks; Torino scale; policy; risk probability

**Highlights**:

* We suggest a color coded scale to communicate the magnitude of global catastrophic and existential risk.
* The scale consists of 6 color codes from white to purple.
* Color estimations are primarily based on probability intervals for human extinction risks in the next 100 years.
* Each risk’s estimation could be adjusted to communicate other aspects of risk’s severity.
* The scale is extended to cover smaller global catastrophic and civilizational collapse risks.

# Introduction

Interest in existential risks (X risks), i.e., risks that could end humanity or drastically curtail its potential, has grown dramatically. In the light of early work such as Bostrom (Bostrom, 2002), J. Leslie (Leslie, 1996), and M. Rees (Rees, 2003), dedicated research groups and organizations have been created, such as the Future of Humanity Institute (FHI), the Global Catastrophic Risk Institute (GCRI), the Future of Life Institute (FLI) and the Centre for the Study of Existential Risk (CSER).

This growing community holds that the near-term risks of human extinction are real, largely because of quick technological development (Rees, 2003). Human extinction is seen as particularly problematic due to the loss of future generations (Beckstead, 2013), potentially changing the fate of the Universe (Tipler, 1997). The way society is structured and the way it behaves probably affects the probability of X risks, so a crucial question is whether there are effective techniques to prevent or mitigate the risks. But even to have this conversation, an instrument is required to estimate the size, probability and urgency of X risks, and to communicate that estimation to society.

One stumbling block to communication is a lack of an exact Х-risk definition (Torres, 2016). The term is not self-evident outside of the community, and it remains open whether X risks should include mass future animal suffering (an example of suffering risks: S risks (Daniel, 2017)) or unmeasurable loss of human potential (Bostrom, 2003), or a risk of the existence of aliens (Baum, Haqq-Misra, & Domagal-Goldman, 2011).

Sometimes the risk’s probability is used as an instrument to communicate its importance, but for many new technological risks we lack methods of determining the probabilities, or we cannot measure them in meaningful ways. Leslie (Leslie, 1996), Bostrom (Bostrom, 2002), Rees (Rees, 2003) and others gave different numerical estimates of the cumulative probability of human extinction. X risks associated with the Large Hadron Collider were presented as very small probabilities (Kent, 2004). Baum and Barrett, discussing the integrated assessment of global catastrophic risk (GCR) suggested measuring the risks through the expected utility of prevention measures (Baum & Barrett, 2017).

There have been some steps towards developing instruments like ours, and we draw on previous efforts. Tonn and Steifel have suggested classifying X risks by their urgency of risk prevention, and we coordinate our estimation of risk severity with this classification (Tonn & Steifel, 2017). Another example is an 80,000-hours attempt to measure all global problems by scale, neglectedness, solvability and personal fit (80,000 hours 2017). The difficulty of estimating small risks with high epistemic uncertainty was analyzed in (Ord, Hillerbrand, & Sandberg, 2010).

So there is a need for an instrument to communicate the importance of X risks. This problem has been solved already for specific risks, one such solution is the Torino scale of asteroid danger. This consists of five color coded levels (from white to red). It seems reasonable to create a similar scale to communicate the magnitude of global catastrophic and X risks, that is our task here.

Vague claims about X risk probabilities are the most common current instrument for communicating X risk magnitudes. This could be misleading, as it poses many technical and ontological difficulties, which we will explore in section 3.

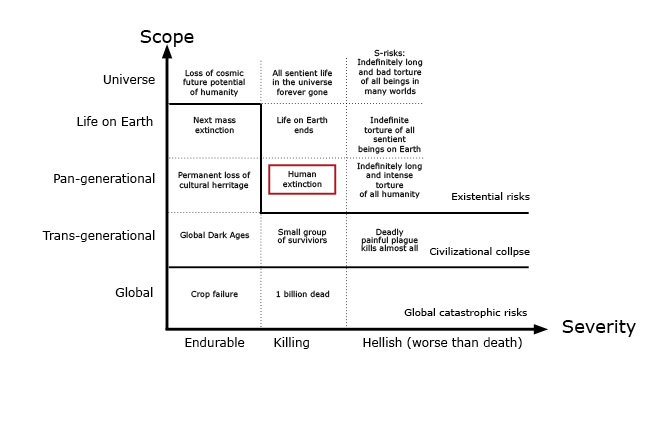
We begin by redefining X risks (section 2), and then discussing what we mean by the probability of such risks (section 3). We then explore existing scales, and finally construct our color coded scale (section 5).

# 2. Human extinction risks as the main type of catastrophic risks and their point of reference

One source of difficulty for defining X risks is ambiguity about what an X risk actually *is.* X risks have been defined in terms of human extinction, a loss of human potential (Bostrom, 2002) or by infinite suffering of human beings or even animals (Daniel, 2017). Note that in several of these cases humans extinction is not necessary (Torres, 2016). Also, the loss of potential and suffering have rather arbitrary and unmeasurable definitions: for instance, aging may be regarded as one of the worst forms of suffering and loss of potential (De Grey & Rae, 2007), and determining the kinds potential humanity may have is very difficult. All this makes “X risks” under the current definition unmeasurable.

We suggest using the term *human extinction risks* to describe a baseline scenario[[1]](#footnote-2): events which lead to *Homo sapiens* becoming extinct. Human extinction risks are the catastrophes in which all human beings die, and no future generations will exist: we include in ‘future generations’ scenarios where humans ‘live on’ via technological interventions, such as via human brain emulations. Other catastrophic risks will be larger or smaller than human extinction, and these will be causally connected with human extinction risks. For example, if all life on Earth went extinct, this would be a worse outcome than “simple” human extinction, as no other civilization could ever appear on Earth. Further, no future Earth-originating civilization would be able to reconstruct part of our archeological heritage.

To present the central role human extinction risks play on our schema, we redrew Bostrom’s classic chart (Bostrom, 2002). We added recently identified S-risks as well as risks associated with civilizational collapse. (Figure 1). Human extinction risks are the natural center of all other catastrophic risks, because smaller risks have some probability of causing human extinction, and larger X risks can be understood as going beyond human extinction. (Surely, civilizational collapse risks are important themselves as we will lose all our cultural value if one happened.) Moreover, governments may allocate from the same resources to fight GCRs and human extinction risks and so one scale for all such risks is valuable.



*Figure 1. Human extinction risk is in the center of all other catastrophic risks (adapted from Bostrom 2003).*

We will now examine the different type of global risks presented in our modified Bostrom’s scale, which will be drawn up below in constructing our color codes.

1. “**Global catastrophic risks”** could kill at least 1 billion people or around 10 per cent of the population, according to Global Challenges Foundations definition (Global Challenges Foundation, 2016). There is great uncertainty about the consequences of such events, and probability that one would trigger civilizational collapse. However, any global catastrophe could start a chain of events resulting in civilizational collapse and even human extinction (Tonn & and MacGregor, 2009). Thus, preventing GCRs also lowers the probability of civilizational collapse and human extinction. Further, failure to recover from a GCR could start a journey down a long path of degradation, leaving our species more susceptible to extinction via natural catastrophes such as supervolcanic eruptions much later in time. Even if there were not loss of civilization, it is possible that the chaos and violence resulting from the catastrophe could damage human values, and then the outcome of AGI might be more negative.
2. “**Civilizational collapse risks**” As most human societies are fairly complex, a true civilizational collapse would require a drastic reduction in human population, and the break-down of connections between surviving populations. Survivors would have to rebuild civilization from scratch, likely losing much technological abilities and knowledge in the process. Hanson (2008) estimated that the minimal human population able to survive is around 100 people. Like X risks, there is little agreement on what is required for civilizational collapse. Clearly, different types and levels of the civilizational collapse are possible (Diamond, 2005) (Meadows, Randers, & Meadows, 2004). For instance, one definition of the collapse of civilization involves, collapse of long distance trade, widespread conflict, and loss of government (Coates, 2009). How such collapses relate to existential risk needs more research.
3. “**Human extinction risks**” are risks that all humans die, and no future generations (in the extended sense mentioned above) will ever exist.
4. “**All life on Earth ends risks”** involve the extinction of all life on earth. As this includes *H. sapiens*, such risks are at the very least on a par with human extinction, but are likely worse as the loss of biodiversity is higher, and (without life arising a second time) no other civilizations, human or otherwise, would be possible on Earth.
5. “**Astronomical scale risks”** include the demise of all civilizations in the affectable universe. This of course includes human extinction, and all life on Earth, and so again are at the very least on a par, and very likely much worse outcomes, than those two.
6. **“S-risks”** include collective infinite suffering (Daniel, 2017). These differ from extinction risks insofar as extinction leads to a lack of existence, whereas this concerns ongoing existence in undesirable circumstances. These also vary in scale and intensity, but are generally out of scope of this work.

Even with a focus squarely on X Risk, global catastrophic risks and civilizational collapse are critically important. This is because there is at least some likelihood that global catastrophic risks increase the probability of human extinction risks—and the more extreme end of civilizational collapses surely would. Before shifting to a discussion of probability appropriate to X risk, we’ll discuss some reasons to link these kinds of risk.

First, global risks may have a fat tail—that is a low probability of high consequences—and the existence of such fat tails strongly depend on the intrinsic uncertainty of global systems (Ćirković, 2012) (Baum, 2015), (Wiener, 2016) (Sandberg & Landry, 2015). This is especially true for risks associated with future world wars, which may include not only nuclear weapons, but weapons incorporating synthetic biology and nanotechnology, different AI technologies, as well as Doomsday blackmail weapons (Kahn, 1959). Another case are the risks associated with climate change, where runaway global warming is a likely fat tail (Obata & Shibata, 2012a), (Goldblatt & Watson, 2012).

Second, global catastrophes could be part of double catastrophe (Baum, Maher, & Haqq-Misra, 2013) or start a chain of catastrophes (Tonn & and MacGregor, 2009), and in this issue (Karieva, 2018). Even if a single catastrophic risk is insufficient to wipe us out, an unhappy coincidence of such events could be sufficient, or under the wrong conditions could trigger a collapse leading to human extinction. Further, global catastrophe could weaken our ability to prepare for other risks.

Luke Oman has estimated the risks of human extinction because of nuclear winter: “The probability I would estimate for the global human population of zero resulting from the 150 Tg of black carbon scenario in our 2007 paper would be in the range of 1 in 10,000 to 1 in 100,000” (Robock, Oman, & Stenchikov, 2007), (Shulman, 2012). Tonn also analyzed chains of events, which could result in human extinction and any global catastrophe may be a start of such chain (Tonn and MacGregor, 2009).

Because this, we suggest that any global catastrophe should be regarded as a possible cause of human extinction risks with no less than 0.01 probability.

Similarly, scenarios involving civilization collapses also plausibly increase the risk of human extinction. If civilization collapses, recovery may be slowed or stopped for a multitude of reasons. For instance, easily accessible mineral and fossil fuel resources might be no longer available, the future climate may be extreme or unstable, we may not regain sufficient social trust after the catastrophe’s horrors, the catastrophe may affect our genetics, a new endemic disease could prevent high population density, and so on. And of course, the smaller populations associated with civilization collapse are more vulnerable to being wiped out by natural events. We estimate that civilization collapse has a 0.1 probability of becoming an existential catastrophe.

In section 4, this discussion will form the basis of our analysis of an X risk’s “severity”, which is the main target of our scale. Before getting there, however, we should first discuss the difficulties of measuring X risks, and related worries regarding probabilities.

# 3. Difficulties of using probability estimates as the communication tool

Plain probability estimates are often used as an instrument to communicate X risks. An example is a claim like “Nuclear war could cause human extinction with probability P”. However, in our view, probability measures are inadequate, both for measuring X risks, and for communicating those risks. This is because of conceptual difficulties (3.1), difficulty in providing meaningful measurements (3.2), the possibility of interaction effects (3.3) and the measurement’s inadequacy for prioritization (3.4) purposes. After presenting these worries, we argue that the *magnitude* of probabilities is a better option, which we use in our tool (3.5).

### 3.1 Difficulties in defining X risk probabilities

Frequentism applies to X risks only with difficulty. One-off events don’t have a frequency, and multiple events are required for frequentist probabilities to apply. Further, on a frequentist reading, claims concerning X risks cannot be falsified. Again, this is because in order to infer from occurrences to probability, multiple instances are required.

Although these conceptual and epistemic difficulties may be analyzed and partly overcome in technical scientific and philosophical literature, they would overcomplicate *a communication tool*.

Also, discussion of X risks sometimes involves weird probabilistic effects. Consider, for example, what (Ćirković, Sandberg, & Bostrom, 2010) call the ‘anthropic shadow’. Because human extinction events entail a lack of humans to observe the event after the fact, we will systematically underestimate the occurrence of such events in an extreme case of survivorship bias (the Doomsday Argument (Tegmark & Bostrom, 2005) is similar)*.*

All of this makes the probabilities attached to X risks extremely difficult to interpret, bad news for an intended communication tool, and stimulates obscure anthropic reasoning. In addition, the subtle features involved in applying frequentism to one-off events, would otherwise tamper with our decision making process.

## 3.2 Data & X Risk

There are little hard data concerning global risks from which probabilities could be extracted. The risk of an asteroid impact is fairly well understood, both due to the historical record, and because scientists can observe particular asteroids and calculate their trajectories. Studies of nuclear winter (Denkenberger & Pearce, 2016), volcanic eruptions, and climate change also provide some risk probability estimates, but are less rigorously supported. In all other cases, especially technological risks, there are many (often contradicting) expert opinions, but little hard data. Those probability calculations which have been carried out are based on speculative assumptions, which carry their own uncertainty.

In the best case, generally, only the order of magnitude of the catastrophe’s probability can be estimated*.* Uncertainty in GCRs is so high, that predictions with high precision are likely to be meaningless. For example, surveys could produce such meaningless over-precision. A survey on human extinction probability gave an estimate of 19 percent in the 21st century (Sandberg & Bostrom, 2008). Such measurements are problematic for communication, because probability estimates of global risks often do not include corresponding confidence intervals(Garrick, 2008). For some catastrophic risks, uncertainty is much larger than for others, because of objective difficulties in their measurement, as well as subjective disagreements between various approaches (especially in the case of climate change, resource depletion, population growth and other politicized areas). As we’ll discuss below, one response is to present probabilities as magnitudes.

## 3.3 Probability density, timing and risks’ interactions

Two more issues with using discrete frequentist probabilities for communicating X risks are related to probability density and the interactions between risks.

For the purpose of responding to the challenges of X risk, the total probability of an event is less useful than the probability density: we want to know not only the probability but the time in which it is measured. This is crucial if policy makers are to prioritize avoidance efforts.

Also, probability estimates of the risks are typically treated separate: interdependence is thus ignored. The total probability of human extinction caused by risk A could strongly depend on the extinction probability caused by risks B and C and also of their timing. (See also double catastrophes discussed by Baum, Maher, & Haqq-Misra, 2013 and the integrated risk assessment project (Baum, 2017).

Further, probability distributions of different risks can have different forms*.*Some risks are linear, others are barrier-like, other logistical. Thus, not all risks can be presented by a single numerical estimate. Exponentially growing risks may be the best way to describe new technologies, such as AI and synthetic biology. Such risks cannot be presented by a single annual probability.

Finally, the probability estimation of a risk depends on whether human extinction is ultimately inevitable. We assume that if humanity becomes an interstellar civilization existing for millions of years, it will escape any near-term extinction risks; the heat death of the universe may be ultimate end, but some think even that is escapable (Dvorsky, 2015). If near-term extinction is inevitable, it is possible to estimate which risks are more probable to cause human extinction (like actuaries do in estimating different causes of death, based in part on the assumption that human death is inevitable). If near-term human extinction is not inevitable, then there is a probability of survival, which is (1- P(all risks)). Such conditioning requires a general model of the future. If extinction is inevitable, the probability of a given risk is just a probability of one way to extinction compared to other ways.

## 3.4 Preventability, prioritizing and relation to the smaller risks

Using bare probability as a communication tool also ignores many important aspects of risks which are substantial for decision makers.

First, a probability estimate does not provide sufficient guidance on how to prioritize prevention efforts. A probability estimate does not say anything about the risk’s relation to other risks, e.g. its urgency.Also, if a risk will take place at a remote time in the future (like the Sun becoming a red giant), there is no reason to spend money on its prevention. Second,a probability estimate does not provide much information about the relation of human extinction risks, and corresponding smaller global catastrophic risks. For example, a nuclear war probability estimate does not disambiguate between chances that it will be a human extinction event, a global catastrophic event, or a regional catastrophe. Third, probability measures do not take preventability into account*.* Hopefully, measures will be taken to try and reduce X risks, and the risks themselves have individual preventability. Generally speaking, it ought to be made clear when probabilities are conditional on whether prevention is attempted or not, and also on the probability of its success.

Probability density, and its relation with cumulative probability could also be tricky, especially as the probability density of most risks is changing in time.

## 3.5 Use of probability orders of magnitude as a communication tool

We recommend using magnitudes of probabilities in communicating about X risk.

One way of overcoming many of the difficulties of using probabilities as communication tool described above is to estimate probabilities with fidelity of one or even two orders of magnitude, and do it over large fixed interval of time, that is the next 100 years (as it the furthest time where meaningful prognoses exist). This order of magnitude estimation will smooth many of the uncertainties described above. Further, prevention actions are typically insensitive in to the exact value of probability. For example, if a given asteroid impact probability is 5% or 25%, needed prevention action will be nearly the same.

For X risks, we suggest using probability intervals of 2 orders of magnitude. Using such intervals will often provide meaningful differences in probability estimates for individual risks. (However, expert estimates sometimes range from “inevitable” to “impossible”, as in AI risks). Large intervals will also accommodate the possibility of one risk overshadowing another, and other uncertainties which arise from the difficulties of defining and measuring X-risks.

This solution is itself inspired by The Torino scale of asteroid danger, which we discuss in more detail below. The Torino scale has five probability intervals, each with a two order of magnitude difference from the next. Further, such intervals can be used to present uncertainty in probability estimation. This uncertainty is often very large for even approximately well-defined asteroid risks. For example, Garrick (Garrick, 2008) estimated that asteroid impacts on the contiguous US with at least 10 000 victims to have expected frequency between once 1: 1900 and 1: 520 000 years with 90 percent confidence. In other words, it used more than 2 orders of magnitude uncertainty.

Of course, there is a lot more to be said about the relationship between X risks and probability—however here we restrict ourselves to those issues most crucial for our purpose, that is, designing a communication tool for X risks.

# 4. Constructing the scale of human extinction risks

## 4.1. Existing scales for different catastrophic risks

In section 2 we established the connection between global catastrophic risks, civilizational collapse risks, human extinction and X risks; we explored the difficulty of the use of probabilities as a communication tool for X risks in section 3; now we can construct the scale to communicate the level of risk of all global catastrophic and X risks.

Our scale is inspired by the Torino scale of asteroid danger which was suggested by professor Richard Binzel (Binzel, 1997). As it only measures the energy of impact, it is not restricted to asteroids but applies to many celestial bodies (comets, for instance). It was first created to communicate the level of risk to the public, because professionals and decision makers have access to all underlying data for the hazardous object.

The Torino scale combines a 5 level color code and 11 level numbered codes. One of the Torino scale’s features is that it connects the size and the probability using diagonal lines, i.e., an event with a bigger size and smaller probability warrants the same level of attention as smaller but more probable events. However, this approach has some difficulties, as was described by (Cox, 2008).

There are several other scales of specific global risks based on similar principles:

1. *Volcanic explosivity index*, VEI, 0-8, (USGS, 2017)
2. *DEFCON* (DEFense readiness CONdition, used by the US military to describe five levels of readiness), from 5 to 1.
3. *“Rio scale” of the Search for Extra-Terrestrial Intelligence* (SETI) – complex scale with three subscales (Almar, 2011).
4. *Palermo scale of asteroid risks* compares the likelihood of the detected potential impactor with the average risk posed by objects of the same size measured both by energy and frequency (NASA, 2017).
5. *San-Marino scale of risks of Messaging to Extra-Terrestrial Intelligence* (METI) (Almar, 2007).

The only more general scale for several global risks is the Doomsday Clock by the Bulletin of the Atomic Scientists, which shows global risks as minutes before midnight. It is oriented towards risks of a nuclear war and climate change and communicates only emotional impact (The Bulletin of the Atomic Scientists, 2017).

## 4.2. The goals of the scale

How good a scale is depends in part on what it is intended to do: who will use it and how will they use it. There are three main groups of people the scale addresses:

**Public**. Simplicity matters: a simple scale is required, similar to the hurricane Saffir-Simpson scale (Schott et al., 2012). This hurricane measuring scale has 5 levels which present rather obscure wind readings as corresponding to the expected damage to houses and thus can help the public make decisions about preparedness and evacuation. In the case of X risks, personal preparedness is not very important, but the public make decisions about which prevention projects to directly support (via donations or crowdfunding) or voting for policymakers who support said projects. Simplicity is necessary to communicate the relative importance of different dangers to a wide variety of non-experts.

**Policymakers**. We intend our scale to help initiate communication of the relative importance of the risks to policymakers. This is particularly important as it appears that policymakers tend to overestimate smaller risks (like asteroid impact risks) and underestimate larger risks (like AI risks) (Bostrom, 2013). Our scale helps to make such comparison possible as it does not depend on the exact nature of the risks. The scale could be applicable to several groups of risks thus allowing comparisons between them, as well as providing a perspective across the whole situation.

**Expert community**. Even a scale of the simplicity we suggest may benefit the expert community. It can act as a basis for comparing different risks by different experts. Given the interdisciplinarity inherent in studying X risk, this common ground is crucial. The scale could facilitate discussion about catastrophes’ probabilities, preventability, prevention costs, interactions, and error margins, as experts from different fields present arguments about the importance of the risks on which they work. Thus it will help to build a common framework for the risk discussions.

## 4.3. Color codes and classification of the needed actions

Tonn and Steifel suggested a six-level classification of actions to prevent X risks (Tonn & Steifel, 2017). They start from “do nothing” and end with “extreme war footing, economy organized around reducing human extinction risk”.

We suggest a scale which is coordinated with Tonn and Steifel’s classification of actions (Table 1), that is our colors correspond to the needed level of action. Also, our colors correspond to typical non-quantifiable ways of the risks description: theoretical, small, medium, serious, high and immediate.

We also add *iconic examples*, which are risks where the probability distribution is known with a higher level of certainty, and thus could be used to communicate the risk’s importance by comparison. Such examples may aid in learning the scale, or be used instead of the scale. For instance, someone could say: “this risk is the same level as asteroid risk”. The iconic risks are marked **bold** in the scale. Iconic examples are also illustrated with the best-known example of that type of event. For example, the best known supervolcanic eruption was the Toba eruption 74,000 years ago (Robock et al., 2009). The Chicxulub impact 66 million years ago is infamous for being connected with the latest major extinction, associated with the non-avian Dinosaur extinction.

The scale presents the total risk of one type of event, without breaking categories down into subrisks. For example, it estimates the total risks of all known and unknown asteroids, but not the risk of any particular asteroid, which is a departure from the Torino scale.

Although the scale is presented using probability intervals, it could be used instead of probabilities if they are completely unknown, but other factors, such as those affecting scope and severity, are known. For example, we might want to communicate that AI catastrophe is a very significant risk, but its exact probability estimation is complicated by large uncertainties. Thus we could agree to represent the risk as *red* despite difficulties of its numerical estimation.

Note that the probability interval (when it is known) for “red” is shorter and is only 1 order of magnitude, as it is needed to represent most serious risks and here we need better resolution ability.

As it is a communication scale, the scientists using it could come to agreement that a particular risk should be estimated higher or lower in this scale. We don’t want to place too many restrictions on how different aspects of a risk’s severity (like preventability or connection with other risks) should affect risks coding, as it should be established in the practical use of the scale. However, we will note two rules:

1. The purple color is reserved to present extreme urgency of the risk

2. The scale is extrapolated from the smaller than extinction risks and larger than extinction risks in Table 2. (This is based on idea that smaller risks have considerable but unknown probability to become human extinction risks, and also on the fact that policy makers may implement similar measures for smaller and larger risks).

## 4.4. Extrapolated version of scale which accounts for the risk size

In Table 2 we extend the scale to include smaller risks like civilization collapse and global catastrophic risks as well as on “larger” risks like life extinction and universe destruction, in accordance with our discussion in section 2. This is necessary because:

1. Smaller risks could become larger extinction risks by starting chains of catastrophic events.
2. The public and policymakers will react similarly to human extinction level catastrophe and to a global catastrophe where there will be some survival: both present similar dangers to personal survival, and in both similar prevention actions are needed.

*Table 1. X risks communication scale: explanation of colors and coordination with Tonn and Steifel’s classification of actions*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Colour coding** | **Explanation**  **of colors** | **Tonn and Steifel’s**  **classification**  **of actions** | **Indicative risk** | **Probability interval** (if known for simple risks**)**  incase of human extinction risks in next 100 years (and medium timing) |
| Purple | **Immediate extreme risks** of human extinction | Extreme war footing, economy organized around reducing human extinction risk | Imminent risk of WW3 beginning (like Cuban missile crisis); | Imminent |
| Red | **High risks** of existential catstophe | Rationing, population control, major command and control regulations | AI control problem | 10-100% |
| Orange | **Serious risks** of global catastrophe, large prevention efforts | Manhattan scale  projects | Full scale nuclear war producing human extinction | 0.1-10%  (once in ~10 000  years event) |
| Yellow | **Medium risks**, which require some prevention activity | Major programs (e.g., carbon tax) and major public investments | Supervolcanic eruption (Toba size event) of extinction level size | 0.001-0.1%  (once in ~1 million  years event) |
| Green | **Small risks**, which require observation | Minor tax incentives, deployment programs; | Asteroid danger (Chicxulub-size event) | 0.00001-0.001%  (once in ~100 million  years event) |
| White | **Theoretical risks** | Do nothing; | Sun becomes red giant | Less than 0.00001%  (once in >1 billion years event) |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Size of the*  *catastrophe:* | All possible civilizations in the visible Universe destroyed  (including humanity) | All life on Earth destroyed (including humanity) | Human extinction | Civilization collapse  (small group of people survives) | Global catastrophe  (above  1 billion victims) |
| *Probability interval*  *in the next 100 years (equivalent timing):* |
| *This row represents immediate near-term danger* |  |  |  |  |  |
| 0.1 - 1 |  |  |  |  | 10% global agricultural shortfall |
| 0.01 - 0.1 |  | **Non-aligned AI** | Synthetic biology risks | **Full scale nuclear war** | Natural pandemic |
| 0.001 - 0.01  (100k-10k years) |  | Nanotech |  |  | **Global warming** |
| 10E-4 - 0.001  (1 mln-100k years) |  |  |  | Super  volcanic eruption |  |
| 10E-5 - 10E-4  (10 mln-1 mln years) |  |  |  | **Asteroid impact** |  |
| 10E-6 - 10E-5  (100 mln-10 mln years) |  |  |  |  |  |
| 10E-7 - 10E-6  (1 bn - 100 million years) |  |  |  |  |  |
| 10E-8 - 10E-7  (10 bn- 1bn years) |  | Collider black hole or strangelet |  |  |  |
| 10E-8  (around 10 billion years) | Collider false vacuum decay | **Sun becomes red giant** |  |  |  |

*Table 2. X risk and global risks combined communication matrix. Probability is estimated for the case if nothing is done to prevent the risk and no other catastrophes happen. For small probabilities we also show corresponding recurrence intervals, but they are only applicable to natural risks, which have background frequencies. In* ***bold*** *are iconic risks for each color code, which are better defined and could be used to represent the type of risk.*

## 4.5. Accessing risks with shorter timeframes than 100 years

In Table 2 above we assessed the risks for the next 100 years. However, without prevention efforts, some risks could approach a probability of 1 in less time: climate change, for instance. We suggest that the urgency of intervening in such cases may be expressed by increasing their color coding.

Moreover, the critical issue is less the timing of risks, but the timing of the prevention measures. Again, although extreme global warming would likely only occur at the end of the 21st century, it is also true that cutting emissions now would ameliorate the situation. We suggest, then, three ranks which incorporate these shorter time-frame risks. Note that the timings relate to implementation of interventions not the timings of the catastrophes.

1) **Now**. This is when a catastrophe has started, or may start in any moment: The Cuban Missile Crisis is an historical example. We reserve purple to represent it.

2) “**Near mode**”. Near mode is roughly the next 5 years. Typically current political problems (as in current relations with North Korea) are understood in near mode. Such problems are appropriately explored in terms of planning and trend expectations. Hanson showed that people are very realistic in “Near mode”, but become speculative and less moral in “Far mode” thinking (Hanson, 2010). Near mode may require one color code increase.

3) **“Next 2-3 decades”**. Many futurists predict a Technological Singularity between 2030-2050: that is around 10-30 years from now (Vinge, 1993), (Kurzweil, 2006). As this mode coincides with an adult’s working life, it may also be called “in personal life time”. In this mode people may expect to personally suffer from a catastrophe, or be personally responsible for incorrect predictions. MIRI recently increased its estimation of the probability that AGI will appear around 2035 (MIRI, 2017), pushing AGI into “next 2-3 decades” mode.

There is a consideration against increasing the color code too much for near-term risks, as that may lead to myopia regarding long-term risks of human extinction. There will always be smaller but more urgent risks, and although these ought to be dealt with, some resources ought to be put towards understanding and mitigating the longer term.

Having said this, in high impact emergency situations, short term overwhelming efforts may help to prevent impending global catastrophe. Examples include the Cuban missile crisis and fighting the recent Ebola pandemic in Western Africa. Such short-term efforts do not necessarily constrain our long-term efforts towards preventing other risks. Thus, short term global catastrophic and larger risks may get a purple rating.

## 4.6. Detailed explanation of risk assessment principles in the color coded scale

In Table 3, we estimate the main global risks, according to the scale suggested in section 4.4.

*Table 3. Detailed explanation of the X risks scale*

|  |  |
| --- | --- |
| Color code | Examples of risks |
| **White** | **Sun becomes red giant.** Although this risk is practically guaranteed, it is very remote indeed.  **Natural false vacuum decay.** Bostrom and Tegmark estimated such events as happening in less than one in 1 billion years, (that is 10-7 in a century) (Tegmark & Bostrom, 2005). Moreover, nothing can be done to prevent it. |
| **Green** | **Gamma-ray bursts.** Earth threatening gamma-ray bursts are extremely rare, and in most cases they will result only in a crop failure due to UV increases. However, a close gamma-ray burst may produce a deadly muon shower which may kill everything up to 3 km in depth (A. Dar, Laor, & N.J, 1997). However, such events could happen less than once in a billion years (10-7 in a century) (Cirković & Vukotića, 2016). Such an event will probably kill all multicellular life on Earth. Dar estimates risks of major extinction events from gamma ray bursts as 1 in 100 mln years (A. Dar, 2001).  **Asteroid impacts.** No dangerous asteroids have been thus far identified, and the background level of global catastrophic impacts is around 1 in a million years (10-4 in a century). Extinction-level impact probability is 10-6 per century. There are several prevention options involving deflecting comets/asteroids. Also, food security could be purchased cheaply (Denkenberger, 2015). However, some uncertainty exists. Some periods involve intense comet bombardment, and if we are in such a time investment in telescopes should be larger (Rampino & Caldeira, 2015).  **High energy accelerator experiments creating false vacuum decay/black hole/strangelet.** Vacuum decay seems to have extremely low probability, far below 10-8 currently. One obvious reason for expecting such events to have very low probability is that similar events happen quite often, and haven’t destroyed everything as yet (Kent, 2004). However, we give this event a higher estimation for two reasons. First, as accelerators become more capable such events might become more likely. Second, the risks are at an astronomical scale: it could affect other civilizations in the universe. Other types of accelerator catastrophes, like mini-black hole or strangelet creation, would only kill Earth life. However, these are more likely, with one estimate being <2E-8 risk from a single facility (the Relativistic Heavy Ion Collider) (Arnon Dar, De Rújula, & Heinz, 1999), which should be coded white. There many unknowns about dangerous experiments (Sandberg & Landry, 2015). Overall, these risks should be monitored, so green is advisable. |
| **Yellow** | **Supervolcanic eruption.** Given historical patterns, the likelihood of living in a century containing a super volcanic eruption is approximately 10-3 (Denkenberger, 2014). However, the chance of human extinction resulting is significantly lower than this. If such an eruption produces global crop failure, it could end current civilization. Conventional wisdom is that there is nothing that could be done to prevent a super volcano from erupting, but some possible preventive measures have been suggested (Denkenberger, this issue). We estimate supervolcanic risks to be higher than asteroid impacts because of the historical record, as they likely nearly finished us off 74 000 ago (Robock et al., 2009).  **Natural pandemic.** A natural pandemic is fairly likely to kill 1% (to an order of magnitude) of the global population during this century, as the Spanish flu did. However, such a pandemic is very unlikely to cause total extinction because lethality is under 100% and some populations are isolated. Between all natural pandemics, emerging pandemic flus have a shorter timespan and need much more attention. Bird flu has a mortality above 0.5 (WHO, 2017) and could produce widespread chaos and possible civilizational collapse if human-to-human transmission starts. Therefore, we estimate 10% probability this century of 10% mortality.  **Global warming triggering global catastrophe**. According to the IPCC anthropogenic global warming may affect billions of people by the end of the 21st century (Parry, 2007), causing heat waves, crop failures and mass migration. Those events, and downstream consequences such as conflicts, could conceivably kill 1 billion people. However, this would only occur for tail risk scenarios which have order of magnitude 1% probability. Having said this, several experts think that methane release from permafrost and similar positive feedback loops may result in runaway global warming with much larger consequences (Obata & Shibata, 2012). |
| **Orange** | **Full-scale nuclear war**. There is roughly 0.02-7% chance per year of accidental full-scale nuclear war between the US and Russia (Barrett, Baum, & Hostetler, 2013). With fairly high probabilities of nuclear winter and civilization collapse given nuclear war, this is order of magnitude 10% this century.  We should also take into consideration that despite reductions in nuclear weapons, a new nuclear arms race is possible in the 21st century. Such a race may include more devastating weapons or cheaper manufacturing methods. Nuclear war could include the creation of large cobalt bombs as doomsday weapons or attacks on nuclear power plants. It could also start a chain of events which result in civilization collapse.  **Nanotechnology risks.** Although molecular manufacturing can be achieved without self-replicating machines (Drexler & Phoenix, 2004), technological fascination with biological systems makes it likely that self-replicating machines will be created. Moreover, catastrophic uses of nanotechnology needn’t be due to accident, but also due to the actions of purposeful malignant agents. Therefore, we estimate the chance of runaway self-replicating machines causing “gray goo” and thus human extinction to be one per cent in this century. There could also be extinction risks from weapons produced by safe exponential molecular manufacturing. See also (Turchin, 2016).  **Artificial pandemic and other risks from synthetic biology.** An artificial multipandemic is a situation in which multiple (even hundreds) of individual viruses created through synthetic biology are released simultaneously either by a terrorist state or as a result of the independent activity of biohackers (Turchin, Green, & Dekenbergern, 2017). Because the capacity to create such a multipandemic could arrive as early as within the next ten to thirty years (as all the needed technologies already exist), it could overshadow future risks, like nanotech and AI, so we give it a higher estimate. There are also other possible risks, connected with synthetic biology, which are widely recognized as serious (Bostrom, 2002).  **Agricultural catastrophe.** There is about a one per cent risk per year of a ten per cent global agricultural shortfall occurring due to a large volcanic eruption, a medium asteroid or comet impact, regional nuclear war, abrupt climate change, or extreme weather causing multiple breadbasket failures (Denkenberger 2016). This could lead to 10% mortality. |
| **Red** | **AI risks.** The risks connected with the possible creation of non-aligned Strong AI are discussed by (Bostrom, 2014), (Yudkowsky, 2008), (Yampolskiy & Fox, 2013) and others. It is widely recognized as the most serious X risk. AI could start an “intelligence explosion wave” through the Universe, which could prevent appearance of the other civilizations before they create their own AI. |
| **Purple** | **Something like the Caribbean crisis in the past, but larger size.**  Currently, there are no known purple risks. If we could be sure that Strong AI will appear in the next 100 years and would probably be negative, it would constitute a purple risk.  Another example would be the creation of a Doomsday weapon that could kill our species with global radiation poisoning (much greater ionizing radiation release than all of the current nuclear weapons) (Kahn, 1959). A further example would be a large incoming asteroid being located, or an extinction level pandemic has begun. These situations require quick and urgent effort on all levels. |

## 4.7. Total risk estimation: red

In addition to using our scale to rank particular risks, we can also use it to estimate our total risk. We think our total risk estimation puts us in the red category.

First, there are around ten main classes and many subtypes of global catastrophes which could happen in the 21st century, and the number is increasing. Second, our technology is growing exponentially and instruments of global destruction are becoming cheaper, especially in biotechnology. Third, the world is divided into many nations and ideologies, with investments of a trillion dollars a year in weapons (Starck, 2005). Fourth, most governments are ignoring X risks, and the total funding of X risk related research is relatively small. Fifth, at the beginning of such rapid increases to our technological powers, we live during a special moment in human history. It is now that X risk prevention measures are most needed. Finally, sixth, most likely there are many unknown risks, which we cannot estimate.

On this basis, we estimate total situation with global catastrophic and X risks as **red**. That is, some kind of global catastrophe will most likely happen in the next 100 years. This means that urgent global efforts are needed to lower the probability of human extinction.

## 4.8. Other possible approaches for the risks communication and classification

It is worth mentioning that there are other approaches which could be used in concert with our color coded system.

1. *Use of the prioritization list to demonstrate which risks are the most important*. For example: 1. AI risks 2. Nuclear war risks 3. Synthetic biology risks.
2. *Create a 3D scale matrix,* using total risk’s probability, magnitude and timing as three axes.
3. *Measure global risks by expected utility* of the prevention efforts, expressed in saved lives or dollars. This approach is used in (Baum & Barrett, 2017).
4. *Doomsday clock*, as a communication tool to express general level of needed emotional impact (The Bulletin of the Atomic Scientists, 2017).

# 5. Conclusion & Further Work

The biggest gains in global safety will be achieved by preventing the highest risks in the proposed scale: near-term risks with the highest probability. Specifically, these are the orange and red risks: AI, synthetic biology, nanotechnology, nuclear war, and agricultural catastrophes.

Less obvious is what to do with lower level risks. Should they be ignored due to falling into the margin of error of estimation of the highest risks? Or should we attempt to prevent all risks proportionally to their perceived level? Another approach is to work on all risks, the thought being that the marginal cost effectiveness of reducing civilization collapse or extinction is equated between the risks. Indeed, putting a small amount of our resources towards researching such low-level risks could be achieved with minimal costs.

Ideally society should provide enough resources to maximally prevent all X risks: both those far and near. If resources are limited, it is better to prevent the near-term risks; if civilization survives longer, hopefully more resources and knowledge will become available to prevent further risks. If risks are imminent, ­and clear to policymakers, they could expand prevention budgets by re-routing funds from other areas.

Global catastrophic and X risks are extremely important, affecting as they do future human and other kinds of life. A risk communication instrument is needed to convey the importance of different risks and thus to prioritize action. We have discussed considerations in constructing a communication scale for global catastrophic and X risks. We estimated several often-discussed types of risks using our scale, and discussed prioritization.

Future work will depend on practical implementation of the scale in the discussion and communication situations and may include upgrade of the set of rules based on this practice.

Future work could include formalizing neglectedness and tractability to refine estimates of cost effectiveness of working on different risks as was attempted by 80,000 Hours (80,000 Hours 2017). Other future work could be a scale for interventions, rather than risks. A more rigorous method of taking into account urgency is estimating return on investment, as was done by (Denkenberger & Pearce, 2016).

# Disclaimer

This article represents the views of the authors, and does not necessarily represent the views of the Global Catastrophic Risk Institute or the Alliance to Feed the Earth in Disasters.

**Acknowledgments**

We would like to thank Alexandra Alexeeva, Seth Baum, Adrian Currie, Anthony Barrett and 2 anonymous reviewers for insightful comments.

**References**:

80,000 Hours. (2017, July 31). How to compare different global problems in terms of impact. Retrieved July 31, 2017, from https://80000hours.org/articles/problem-framework/

Almar, I. (2007). The San Marino Scale: A new analytical tool for assessing transmission risk. *Acta Astronautica, Volume 60, Issue 1, P. 57-59.*

Almar, I. (2011). SETI and Astrobiology: The Rio Scale and the London Scale. *Acta Astronautica. Science Direct*, *69*, 899–904. https://doi.org/doi:10.1016/j.actaastro.2011.05.036

Barrett, A. M., Baum, S. D., & Hostetler, and K. (2013). Analyzing and reducing the risks of inadvertent nuclear war between the United States and Russia. *Science & Global Security*, *21*(2), 106–133.

Baum, S. D. (2015). Risk and resilience for unknown, unquantifiable, systemic, and unlikely/catastrophic threats. *Environment Systems and Decisions*, *35*(2), 229–236.

Baum, S. D. (2017). Integrated Assessment Project | Global Catastrophic Risk Institute. Retrieved from http://gcrinstitute.org/integrated-assessment/

Baum, S. D., & Barrett, A. M. (2017). *Towards an Integrated Assessment of Global Catastrophic Risk* (SSRN Scholarly Paper No. ID 3046816). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=3046816

Baum, S. D., Haqq-Misra, J. D., & Domagal-Goldman, S. D. (2011). Would contact with extraterrestrials benefit or harm humanity? A scenario analysis. *Acta Astronautica*, *68*(11), 2114–2129.

Baum, S. D., Maher, T. M., & Haqq-Misra, J. (2013). Double catastrophe: intermittent stratospheric geoengineering induced by societal collapse. *Environment Systems & Decisions*, *33*(1), 168–180.

Baum, S. D., Maher, T. M., & Haqq-Misra, J. (2013). Double catastrophe: intermittent stratospheric geoengineering induced by societal collapse. *Environment Systems & Decisions*, *33*(1), 168–180.

Beckstead, N. (2013). *On the overwhelming importance of shaping the far future*. New Brunswick, NJ: Department of Philosophy, Rutgers university.

Binzel, R. P. (1997). A Near-Earth Object Hazard Index. Annals of the. *New York Academy of Sciences*, *822*, 545–551. https://doi.org/doi:10.1111/j.1749-6632.1997.tb48366.x

Bostrom, N. (2002). Existential risks: Analyzing Human Extinction Scenarios and Related Hazards. *Journal of Evolution and Technology, Vol. 9, No. 1 (2002).*

Bostrom, N. (2003). Astronomical waste: The opportunity cost of delayed technological development. *Utilitas*, *15*(3), 308–314.

Bostrom, N. (2013). Existential risk prevention as global priority. *Global Policy*, *4*, 15–31.

Bostrom, N. (2014). *Superintelligence*. Oxford: Oxford University Press.

Ćirković, M. M. (2012). Small theories and large risks—is risk analysis relevant for epistemology? *Risk Analysis*, *32*(11), 1994–2004.

Ćirković, M. M., Sandberg, A., & Bostrom, N. (2010). Anthropic shadow: observation selection effects and human extinction risks. *Risk Analysis, Vol. 30, No. 10, 2010*.

Cirković, M. M., & Vukotića, B. (2016). Long-term prospects: Mitigation of supernova and gamma-ray burst threat to intelligent beings. *Acta Astronautica*, *129*, 438–446.

Coates, J. F. (2009). Risks and threats to civilization, humankind, and the earth. *Futures*, *41*(10), 694–705.

Cox, L. A. (2008). What’s wrong with risk matrices? *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *28*(2), 497–512. https://doi.org/10.1111/j.1539-6924.2008.01030.x

Daniel, M. (2017). S-risks: Why they are the worst existential risks, and how to prevent them (EAG Boston 2017). Foundational research institute.

Dar, A. (2001). The threat to life from Eta Carinae and gamma ray bursts. *Astrophysics and Gamma Ray Physics in Space (Eds. A. Morselli and P. Picozza), Frascati Physics Series Vol. XXIV (2002), Pp. 513-523*. Retrieved from https://arxiv.org/abs/astro-ph/0110162

Dar, A., De Rújula, A., & Heinz, U. (1999). Will relativistic heavy-ion colliders destroy our planet? *Physics Letters B*, *470*(1), 142–148.

Dar, A., Laor, A., & N.J, S. (1997). Life extinctions by cosmic ray jets. *Physical Review Letters*, *80*(26). Retrieved from https://arxiv.org/pdf/astro-ph/9705008v1.pdf

De Grey, A., & Rae, M. (2007). *Ending aging: The rejuvenation breakthroughs that could reverse human aging in our lifetime*. St. Martin’s Press.

Denkenberger, D., & Pearce, J. M. (2014). *Feeding everyone no matter what: managing food security after global catastrophe*. Academic Press.

Denkenberger, D., & Pearce, J. M. (2015). Feeding everyone: Solving the food crisis in event of global catastrophes that kill crops or obscure the sun. *Futures*, *72*, 57–68.

Denkenberger, D., & Pearce, J. M. (2016). Cost-Effectiveness of Interventions for Alternate Food to Address Agricultural Catastrophes Globally. *International Journal of Disaster Risk Science*, *7*(3), 205–215.

Diamond, J. (2005). *Collapse: How societies choose to fail or succeed*. Penguin.

Drexler, E., & Phoenix, C. (2004). Safe exponential manufacturing. *Nanotechnology*, *15*, 869.

Dvorsky, G. (2015). Will Our Descendants Survive the Destruction of the Universe? IO9. Retrieved from https://io9.gizmodo.com/will-our-descendants-survive-the-destruction-of-the-uni-1744169933

Garrick, B. J. (2008). *Quantifying and Controlling Catastrophic Risks*. Academic Press.

Global Challenges Foundation. (2016). Global Catastrophic Risks 2016. Retrieved from https://api.globalchallenges.org/static/reports/Global-Catastrophic-Risk-Annual-Report-2016.pdf

Goldblatt, C., & Watson, A. J. (2012). The runaway greenhouse: implications for future climate change, geoengineering and planetary atmospheres. *Phil. Trans. R. Soc. A*, *370*(1974), 4197–4216. https://doi.org/10.1098/rsta.2012.0004

Hanson, R. (2010). Near-Far Summary. Retrieved from http://www.overcomingbias.com/2010/06/near-far-summary.html

Kahn, H. (1959). *On thermonuclear war*. Princeton University Press.

Karieva, P. (2018). Existential Risk due to Ecosystem Collapse: Nature Strikes Back. *Futures*, *forthcoming*.

Kent, A. (2004). A critical look at risk assessments for global catastrophes. *Risk Analysis*, *24*(1), 157–68.

Kurzweil, R. (2006). *Singularity is Near*. Viking.

Leslie, J. (1996). *The End of the World: The Science and Ethics of Human Extinction*. Psychology Press.

Matheny, J. G. (2007). Reducing the risk of human extinction. *Risk Analysis*, *27*, 1335–1344. https://doi.org/doi:10

Meadows, D., Randers, J., & Meadows, and D. (2004). *Limits to growth: The 30-year update*. Chelsea Green Publishing.

MIRI. (2017). 2017 Updates and Strategy. Retrieved from https://intelligence.org/2017/04/30/2017-updates-and-strategy/

NASA. (2017). THE PALERMO TECHNICAL IMPACT HAZARD SCALE. Retrieved from http://neo.jpl.nasa.gov/risk/doc/palermo.html

Obata, A., & Shibata, K. (2012a). Damage of land biosphere due to intense warming by 1000-fold rapid increase in atmospheric methane: Estimation with a climate–carbon cycle model. *Journal of Climate*, *25*(24), 8524–8541.

Obata, A., & Shibata, K. (2012b). Damage of land biosphere due to intense warming by 1000-fold rapid increase in atmospheric methane: Estimation with a climate–carbon cycle model. *Journal of Climate*, *25*(24), 8524–8541.

Ord, T., Hillerbrand, R., & Sandberg, A. (2010). Probing the improbable: methodological challenges for risks with low probabilities and high stakes. *Journal of Risk Research*, *13*(2), 191–205.

Parry, M. L. (2007). *Climate change 2007-impacts, adaptation and vulnerability: Working group II contribution to the fourth assessment report of the IPCC* (Vol. 4). Cambridge University Press.

Rampino, M. R., & Caldeira, K. (2015). Periodic impact cratering and extinction events over the last 260 million years. *Monthly Notices of the Royal Astronomical Society*, *454*(4), 3480. https://doi.org/10.1093/mnras/stv2088

Rees, M. (2003). *Our Final Century*. Heinemann.

Robock, A., Ammann, C. M., Oman, L., Shindell, D., Levis, S., & Stenchikov, G. (2009). Did the Toba volcanic eruption of∼ 74 ka BP produce widespread glaciation? *Journal of Geophysical Research: Atmospheres*, *114*(D10).

Robock, A., Oman, L., & Stenchikov, G. L. (2007). Nuclear winter revisited with a modern climate model and current nuclear arsenals: Still catastrophic consequences. *J. Geophys. Res. Atmos.*, *112*(D13), 1984–2012.

Sandberg, A., & Bostrom, N. (2008). Global catastrophic risks survey. *Civil Wars*, *98*(30), 4.

Sandberg, A., & Landry, F. (2015). Putting out the dark fire: constraining speculative physics disasters.

Schott, T., Landsea, C., Hafele, G., Lorens, J., Taylor, A., Thurm, H., … Zaleski, W. (2012). The Saffir-Simpson hurricane wind scale. *NOAA/National Weather Service [Internet]*, 1–4.

Shulman, C. (2012). Overcoming Bias : Nuclear winter and human extinction: Q&A with Luke Oman. Retrieved from http://www.overcomingbias.com/2012/11/nuclear-winter-and-human-extinction-qa-with-luke-oman.html

Starck, P. (2005). World Military Spending Topped $1 Trillion in 2004. *Common Dreams News Center*, *7*, 607–3.

Tegmark, M., & Bostrom, N. (2005). *How unlikely is a doomsday catastrophe?* (Vol. 438). Nature, 1: 438-754. Retrieved from https://arxiv.org/abs/astro-ph/0512204

The Bulletin of the Atomic Scientists. (2017). Doomsday Clock. Retrieved from http://thebulletin.org/timeline

Tipler, F. . (1997). *The physics of immortality: Modern cosmology, God, and the resurrection*.

Tonn, B., & and MacGregor, D. (2009). A singular chain of events. *Futures*, *41*(10), 706–714.

Tonn, B., & Steifel, D. (2017). Human extinction risk and uncertainty: Assessing conditions for action. *Futures, Accepted Manuscript*.

Torres, P. (2016). Problems with defining an existential risk. *IEET*. Retrieved from https://ieet.org/index.php/IEET2/more/torres20150121

Turchin, A. (2016). The map of nanotech risks. Retrieved from http://lesswrong.com/lw/ncs/the\_map\_of\_nanotech\_global\_catastrophic\_risks/

Turchin, A., Green, B., & Dekenbergern, D. (2017). Multiple Simultaneous Pandemics as Most Dangerous Global Catastrophic Risk Connected with Bioweapons and Synthetic Biology. *Forthcoming*.

USGS. (2017). The Volcanic Explosivity Index (VEI). Retrieved from https://volcanoes.usgs.gov/vsc/glossary/vei.html

Vinge, V. (1993). Technological singularity (pp. 30–31). Presented at the VISION-21 Symposium sponsored by NASA Lewis Research Center and the Ohio Aerospace Institute, VISION-21.

WHO. (2017). WHO | FAQs: H5N1 influenza. Retrieved from http://www.who.int/influenza/human\_animal\_interface/avian\_influenza/h5n1\_research/faqs/en/

Wiener, J. B. (2016). The tragedy of the uncommons: On the politics of apocalypse. *Global Policy*, *7*(S1), 67–80.

Yampolskiy, R., & Fox, J. (2013). Safety Engineering for Artificial General Intelligence. Retrieved from https://intelligence.org/files/SafetyEngineering.pdf

Yudkowsky, E. (2008). *Artificial Intelligence as a Positive and Negative Factor in Global Risk, in Global Catastrophic Risks*. (M. M. Cirkovic & N. Bostrom, Eds.). Oxford University Press: Oxford, UK.

1. Several other authors also used term of “human extinction risks” to unambiguously describe this type of global catastrophes (Leslie, 1996; Matheny, 2007). [↑](#footnote-ref-2)