Preprint: Denkenberger, D., Sandberg, A., Tieman, R.J. and Pearce, J.M., Long term cost-effectiveness of resilient foods for global catastrophes compared to artificial general intelligence safety. *International Journal of Disaster Risk Reduction* 73, 2022, 102798. https://doi.org/10.1016/j.ijdrr.2022.102798

Long term cost-effectiveness of resilient foods for global catastrophes compared to artificial general intelligence

David Denkenberger, Anders Sandberg, Ross John Tieman, Joshua M. Pearce

ABSTRACT

Global agricultural catastrophes, which include nuclear winter and abrupt climate change, could have long-term consequences on humanity such as the collapse and nonrecovery of civilization. Using Monte Carlo (probabilistic) models, we analyze the long-term cost-effectiveness of resilient foods (alternative foods) - roughly those independent of sunlight such as mushrooms. One version of the model populated partly by a survey of global catastrophic risk researchers finds the confidence that resilient foods is more cost effective than artificial general intelligence safety is ~86% and ~99% for the 100 millionth dollar spent on resilient foods at the margin now, respectively. Another version of the model based on one of the authors produced ~95% and ~99% confidence, respectively. Considering uncertainty represented within our models, our result is robust: reverting the conclusion required simultaneously changing the 3-5 most important parameters to the pessimistic ends. However, as predicting the long-run trajectory of human civilization is extremely difficult, and model and theory uncertainties are very large, this significantly reduces our overall confidence. Because the agricultural catastrophes could happen immediately and because existing expertise relevant to resilient foods could be co-opted by charitable giving, it is likely optimal to spend most of the money for resilient foods in the next few years. Both cause areas generally save expected current lives inexpensively and should attract greater investment.

1. INTRODUCTION

Several global catastrophes have the potential to collapse global agricultural by blocking the sun. Arguably the greatest of these sun blocking catastrophes is full-scale nuclear war between the U.S. and Russia, which would burn many cities and release enough smoke to block the sun for 5-10 years [1,2]. Super volcanic eruption [3], or a large asteroid/comet [4,5], could also have a major impact global agriculture but are lower probability than full scale nuclear war, estimated at ~1% per year [6–8].

A significant reduction in global temperatures is anticipated to result from the reduction in sunlight with corresponding impacts on agricultural productivity during such an event. Several pieces of evidence indicate impacts will be severe. Considering nuclear conflicts, a limited regional conflict between Pakistan and India, in which ~ 100 x 15 kt (thousand tons of TNT equivalent) atomic bombs would be detonated indicated ~1.5°C global cooling [9]. Recent modelling of full scale nuclear exchanges between the U.S. and Russia demonstrated pyro cumulonimbus injections of smoke into the stratosphere from forest fires could cause sun blocking and corresponding nuclear winter for ~5 years [1]. Climatic cooling and resulting ecological impacts that followed the ~74 Ka Toba super eruption provide another piece of evidence though it is unclear how this translates to impacts on present day agriculture [10–12]. Prolonged climatic cooling and glaciation ~1000 years long appear unlikely; however, cooling of ~10 K over a period of years is plausible [11] and would have catastrophic impacts on agriculture.

Collapse of global agriculture can be thought of as a 100% order of magnitude reduction in output, while less severe scenarios can be thought of as 10% disruptions. The geometric mean between orders of magnitude provides the most appropriate delineation of impacts ranges, giving the following agricultural impact scenarios: 0 - 3%, relatively commonplace and typically easily absorbed by the current food system; 3-30%, greater than any disruption in the past half century and likely to cause global harm [13], and 30-100% unprecedented shock with significant potential to cause civilizational collapse.

The collapse of agriculture is of concern due to its critical function in sustaining civilization and the various beneficial functions and services civilization provides humanity [14] (e.g. food, shelter, goods, medicine, education, global communication, and advanced technologies). These benefits, along with many more, would be forfeited in the event of civilizational collapse. One definition of the collapse of civilization involves short-term focus, collapse of long distance trade, widespread conflict, and loss of government [15]. For those that subscribe to long termism [16], the view that the future should have a near zero discount rate, the unrecoverable loss of civilization appears even more impactful.

If civilization were lost it may not be easily recovered for various reasons: Readily accessible fossil fuels and minerals are exhausted, there might not be the stable climate of the last 10,000 years, or trust or IQ might be lost permanently because of the trauma and genetic selection of the catastrophe. Another route to far future impact is the trauma associated with the catastrophe making future catastrophes more likely, such as global totalitarianism [17]. A further route is worse values caused by the catastrophe could be locked in by artificial general intelligence (AGI)[18]. If the loss of civilization persists long enough, a natural catastrophe could cause the extinction of humanity. Catastrophes such as nuclear winter, super volcanic eruption, or a large asteroid/comet impact could directly cause human extinction [15][1].

Preventing the risk of the collapse of civilization from nuclear war can be achieved through several risk mitigation strategies. The obvious risk mitigation intervention is prevention of nuclear war, which would be the best outcome. However, it is not neglected, as it has been worked on for many decades and is currently funded at billions of dollars per year quality adjusted [20]. As the largest problem in sunblocking catastrophes is that of food supply [21], the next most obvious solution is storing food. Storing five to 10 years of food, however, for everyone globally is very expensive [22], takes too long for sufficient storage to be achieved for near-term catastrophes, and would result in additional millions being malnourished in the short term if stored quickly. In these scenarios, the majority of infrastructure and industry would still be functioning, unlike in scenarios

[1] Though there were concerns that full scale nuclear war would kill everyone with radioactivity, Hiroshima and Nagasaki were continuously inhabited [19]. It turns out that most of the radioactivity is rained out within a few days. One possible mechanism for extinction would be that the hunter gatherers would die out because they do not have food storage. And people in developed countries would have food storage, but might not be able to navigate the path back to being hunter gatherers.

that disrupt electricity such as a solar storm [23–25]. Therefore, one potential mechanism to feed people would be through the deployment of resilient foods (alternative foods)[2].

Resilient foods can be described as novel or adapted methods of food production that are less dependent in sunlight. For instance several resilient foods use alternative energy sources to sunlight such as biomass (e.g. mushrooms and cellulose-digesting mammals/insects, cellulosic sugar)[26], fossil fuels (e.g. methane digesting single cell protein) or other energy sources (e.g. hydrogen oxidizing bacteria, microbial electrosynthesis, and direct chemical (non-biological) synthesis of food) [27–30]. Another example is macroalgae which has the potential to produce significant calories and nutrients in low sunlight and temperatures independently from land and fresh water [31]. Mass production of simple greenhouses in the tropics could as be viable despite lower sunlight [32]. These foods could feed everyone several times over in terms of calories [21], and micronutrients would be adequate as well [33,34]. Developing resilience to sun blocking catastrophes through preparedness planning, research and development appears technically feasible and cost effective, saving expected lives in the present generation in the US for $1 to $20,000 [35]. Resilient foods cannot save the lives of people suffering the direct effects of the catastrophe, for nuclear exchange this consists of the blasts, firestorms and short-term radiation. But since, according to one estimate, about 90% of the remaining population [36] would die with the world's current ~half a year of food storage if the sun were blocked for 5 years, resilient foods could solve ~90% of the most impactful aspect of the catastrophe[3].

Resilient foods could also mitigate the impacts of various catastrophes that could occur in the 21st century that would cause a 10% reduction in global agriculture, examples include: abrupt regional climate change (10°C in a decade) [37], super crop pathogen [38], superweeds [39], super crop pest (animal, e.g. insect) [40], coincident extreme weather, resulting in multiple breadbasket failures (Bailey et al., 2015), slow climate change that is extreme (>~5°C) [41] or smaller or regional nuclear war (for example, India-Pakistan) [42]. Though it would be technically straightforward to reduce food consumption by 10% by making less food go to waste, animals, and biofuels, the prices would go so high that those in poverty may not be able to afford food [43] One study estimated 500 million expected lives would be lost in such a catastrophe [35] Expected lives could be saved globally by preparing for these catastrophes for $0.20 to $400 for only 10% global agricultural shortfalls [35]; cost effectiveness would increase if sun blocking scenarios were considered.

Resilient foods demonstrate potential to reduce the significant impacts of agricultural catastrophes; however, current funding appears disproportionately low as compared to mitigation measures of other global catastrophic risks (GCR) or existential risks. For instance, artificial general intelligence (AGI) is considered a major existential risk [44]. The artificial intelligence that has been available until very recently is narrow AI, i.e. it can only do a specific task, such as playing *Jeopardy!* [45]. Already, AI has become more general with GPT-3 writing, coding, and generating images [46]. There are concerns that as AI systems become more capable, AGI will eventually be

[2] Resilient foods are also referred to as alternative foods in the literature. Resilient foods is used here to better highlight the useful characteristic of the discussed food production methods, i.e., resilience to stressors that would cause traditional agriculture to fail, e.g., loss of sunlight.

[3] This estimate assumed no subsequent conflict, no charity/aid, no migration, but that there would be trade and a single global food price.

achieved [18]. Since AGI could perform all human tasks at least as well as humans, this would include programming AGI. This enables recursive self-improvement, so there could be an intelligence explosion [47]. Given that the goals of the intelligence are essentially arbitrary [48] and could be pursued with great power, this implies a potentially serious risk [18]. AGI safety has been a top priority in the existential risk community[4] [51]. Though there is uncertainty in when and how AGI will be developed, there are concrete actions that can be taken now to mitigate the risk [52].

The objective of this paper is to compare the cost effectiveness of resilient foods with AGI safety to assess whether resilient foods should also be a top priority. Comparisons to other risks, such as asteroids [53], climate change [54] and pandemics [55], are also candidates. However, these are generally regarded by the existential risk community as lower priority, so if resilient foods were more cost effective than AGI safety, they could be the highest priority. A secondary objective is to explore the value and limitations of relative long-term cost effectiveness analysis as a prioritization tool for disaster risk mitigation measures in order to improve decision making. Inclusion of 10% agricultural loss scenarios also provides an opportunity to explore flow through impacts of less extreme, but high likelihood agricultural loss scenarios building on previous work.

2. METHODS

2.1 Overview

Given the large uncertainties in parameters, cost effectiveness is calculated using a Monte Carlo model, producing a probability distribution of expected cost effectiveness. Probabilistic uncertainty analysis is in wide use in insurance, decision-support and cost-effectiveness modelling [56]. In these models, uncertain parameters are represented by samples drawn from given distributions that are combined into output samples that form an empirical distribution.

The models consist of two *independent* submodels: a resilient foods submodel estimating the risk and mitigation costs of food shortfalls, and an AGI risk submodel estimating risk and mitigation costs of AGI extinction scenarios. These two submodels then allow us to estimate the relative ratio of cost effectiveness. Models assume there is a finite upper limit of value achievable by humanity[5] at some point in the future[6] (referred to as potential of humanity or far future value); this limit is consistent across submodels [59].

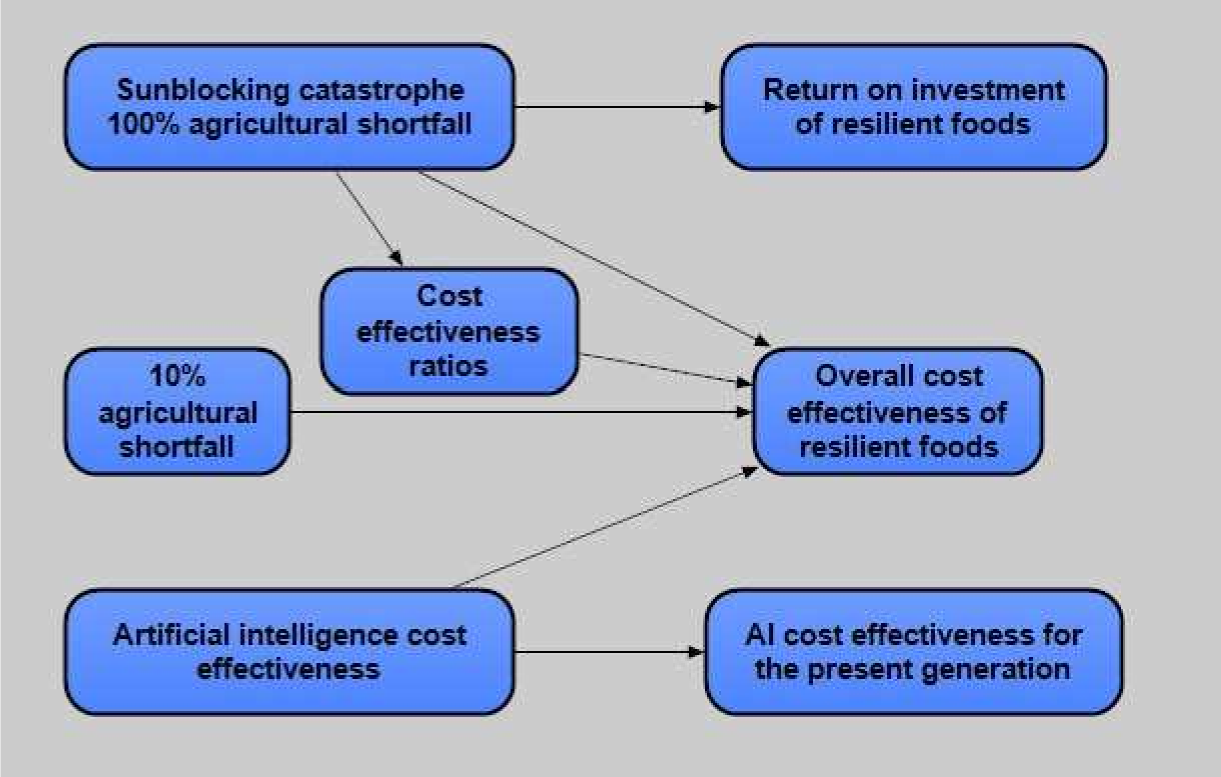
Monte Carlo estimation was selected for flexibility because the probability distributions for various parameters do not conveniently come in a form that provides analytically tractable combinations. It also allows exploring different models and scenarios, examining parameter sensitivity (probabilistic sensitivity analysis).

[4] The existential risk community consist of various organisations, individuals and philanthropic groups who focus on reducing the risk of humanity going extinct through natural or anthropogenic causes [49,50].

[5] This includes ethically relevant agents descending from humanity e.g. AI's, post humans, etc.

[6] The Oxford Prioritisation Project – Machine Intelligence Research Institute Cost-effectiveness Model, which the AGI submodel is based on, attempts to quantify expected value of far future, in human-equivalent well-being-adjusted life-years (HEWALYs)[57,58]

The models were first implemented in open source software called Guesstimate[7], and they are available online. However, for more powerful analysis and plotting, the models were also implemented on the software Analytica 5.2.9. Combining the uncertainties in all the inputs was performed with a Median Latin Hypercube analysis (analogous to Monte Carlo, but better performing [60]) with the highest sample of 32,000 (run time on a personal computer was seconds). The results from the two implementations agreed within uncertainties due to finite number of samples, giving higher confidence in the results.
The two models have identical structure, but differ slightly on choice of input distributions and parameter values. The first model, hence force referred to as the S*urvey model* and abbreviated to S model*,* has inputs defined by Denkenberger and incorporates GCR survey results. The second model, henceforth referred to as the *Expert model* and abbreviated to E model has inputs defined by Sandberg [61].[8] Figures 1 to 4 illustrate the interrelationships of the nodes for S model


Sunblocking catastrophe 100% agricultural shortfall
Return on investment of resilient foods
Cost effectiveness ratios
10% agricultural shortfall
Overall cost effectiveness of resilient foods
Artificial intelligence cost effectiveness
AI cost effectiveness for the present generation

[7] One can change numbers in the viewing mode of the model to see how outputs change, but they will not save. If one wants to save, one can make a copy of the model.

[8] Lognormal results in the median being the geometric mean of the ends (multiply the 5th and 95th percentiles and take the square root). Note that the mean is generally much higher than the median.

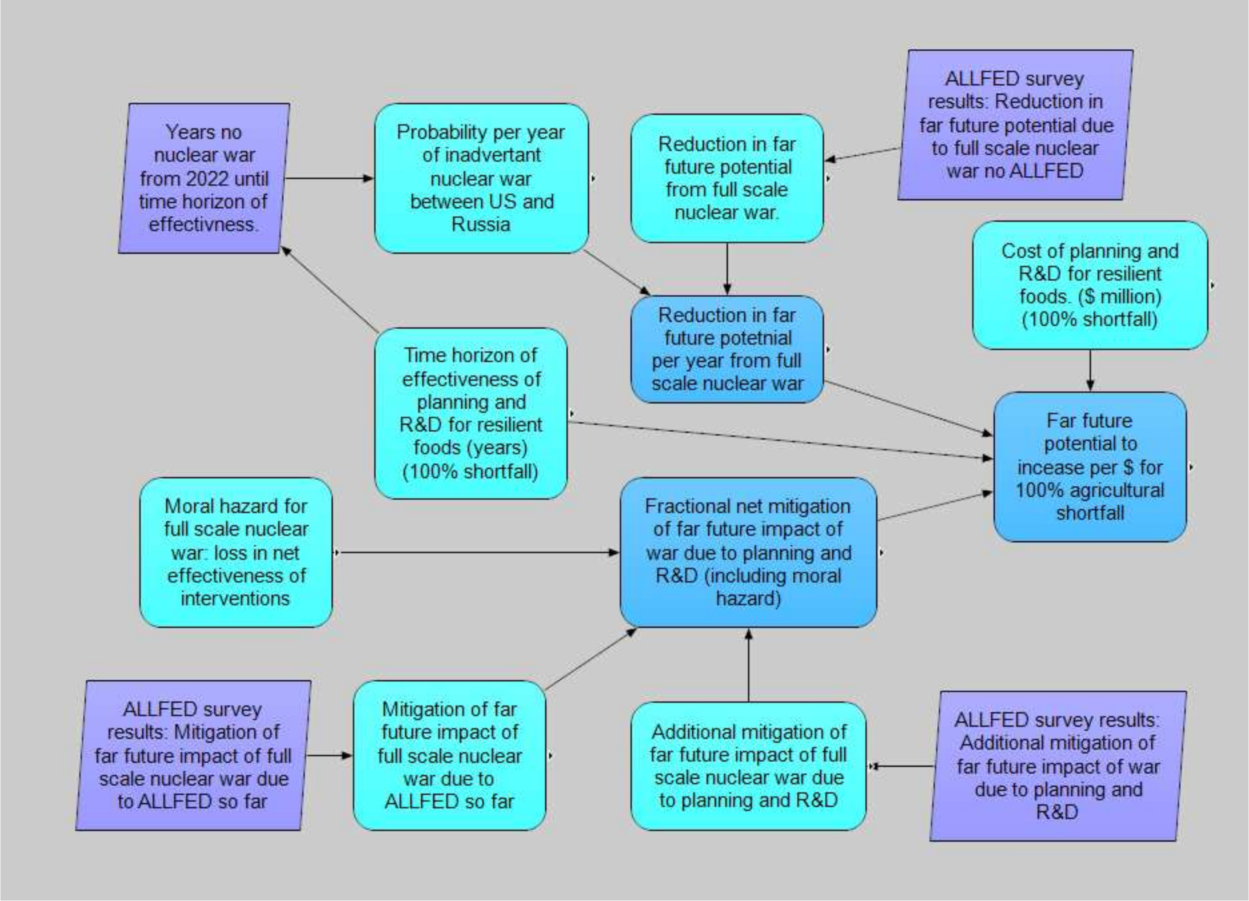
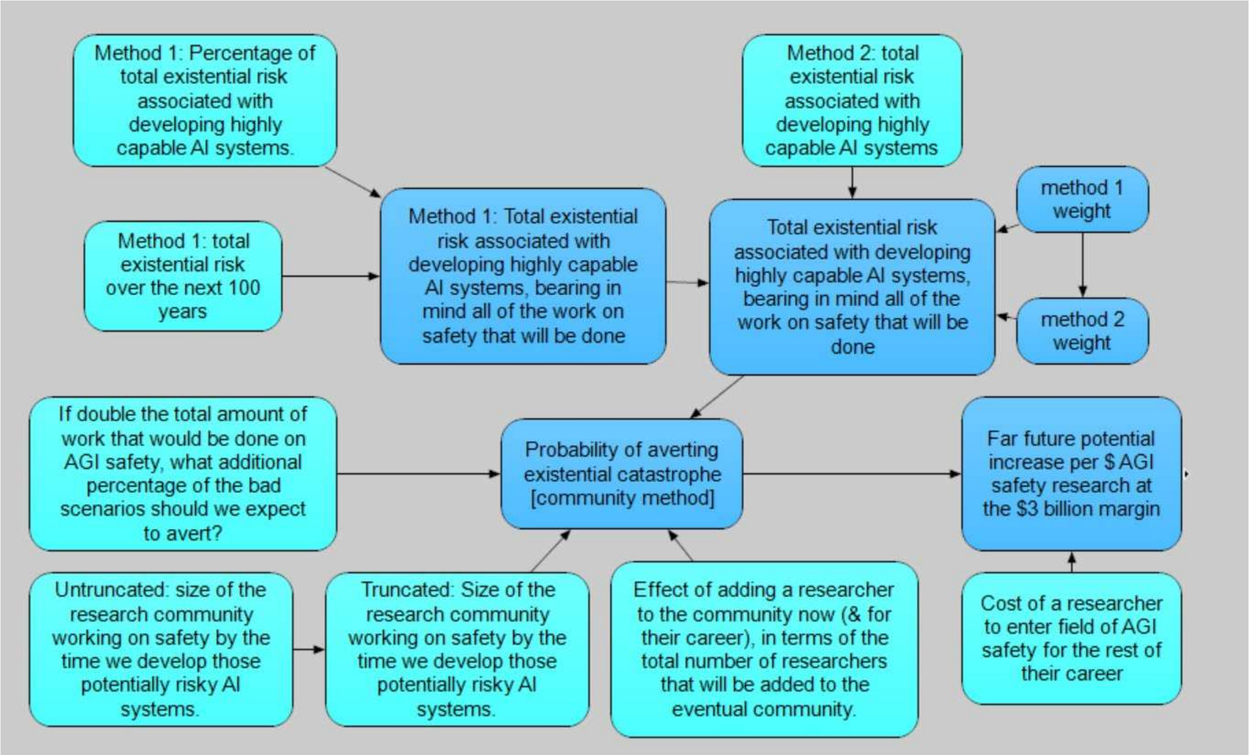

Years no nuclear war from 2022 until time horizon of effectivness.
Probability per year of inadvertant nuclear war between US and Russia
Reduction in far future potential from full scale nuclear war.
ALLFED survey results: Reduction in far future potential due to full scale nuclear war no ALLFED
Cost of planning and R&D for resilient foods. ($ million) (100% shortfall)
Reduction in far future potetnial per year from full scale nuclear war
Time horizon of effectiveness of planning and R&D for resilient foods (years) (100% shortfall)
Far future potential to incease per $ for 100% agricultural shortfall
Moral hazard for full scale nuclear war: loss in net effectiveness of interventions
Fractional net mitigation of far future impact of war due to planning and R&D (including moral hazard)
ALLFED survey results: Mitigation of far future impact of full scale nuclear war due to ALLFED so far
Mitigation of far future impact of full scale nuclear war due to ALLFED so far
Additional mitigation of far future impact of full scale nuclear war due to planning and R&D
ALLFED survey results: Additional mitigation of far future impact of war due to planning and R&D

Figure 2. S model sunblocking catastrophe 100% agricultural shortfall submodel; 10% agricultural shortfall submodel has nearly identical causal structure. The colouring (and shape) of nodes corresponds to type of node, turquoise nodes are chance nodes containing probabilistic input distributions, purple nodes are index nodes and light blue are variable nodes.


Method 1: Percentage of total existential risk associated with developing highly capable AI systems.
Method 2: total existential risk associated with developing highly capable AI systems
Method 1: total existential risk over the next 100 years
Method 1: Total existential risk associated with developing highly capable AI systems, bearing in mind all of the work on safety that will be done
Total existential risk associated with developing highly capable AI systems, bearing in mind all of the work on safety that will be done
method 1 weight
method 2 weight
If double the total amount of work that would be done on AGI safety, what additional percentage of the bad scenarios should we expect to avert?
Probability of averting existential catastrophe [community method]
Far future potential increase per $ AGI safety research at the $3 billion margin
Untruncated: size of the research community working on safety by the time we develop those potentially risky AI systems.
Truncated: Size of the research community working on safety by the time we develop those potentially risky AI systems.
Effect of adding a researcher to the community now (& for their career), in terms of the total number of researchers that will be added to the eventual community.
Cost of a researcher to enter field of AGI safety for the rest of their career

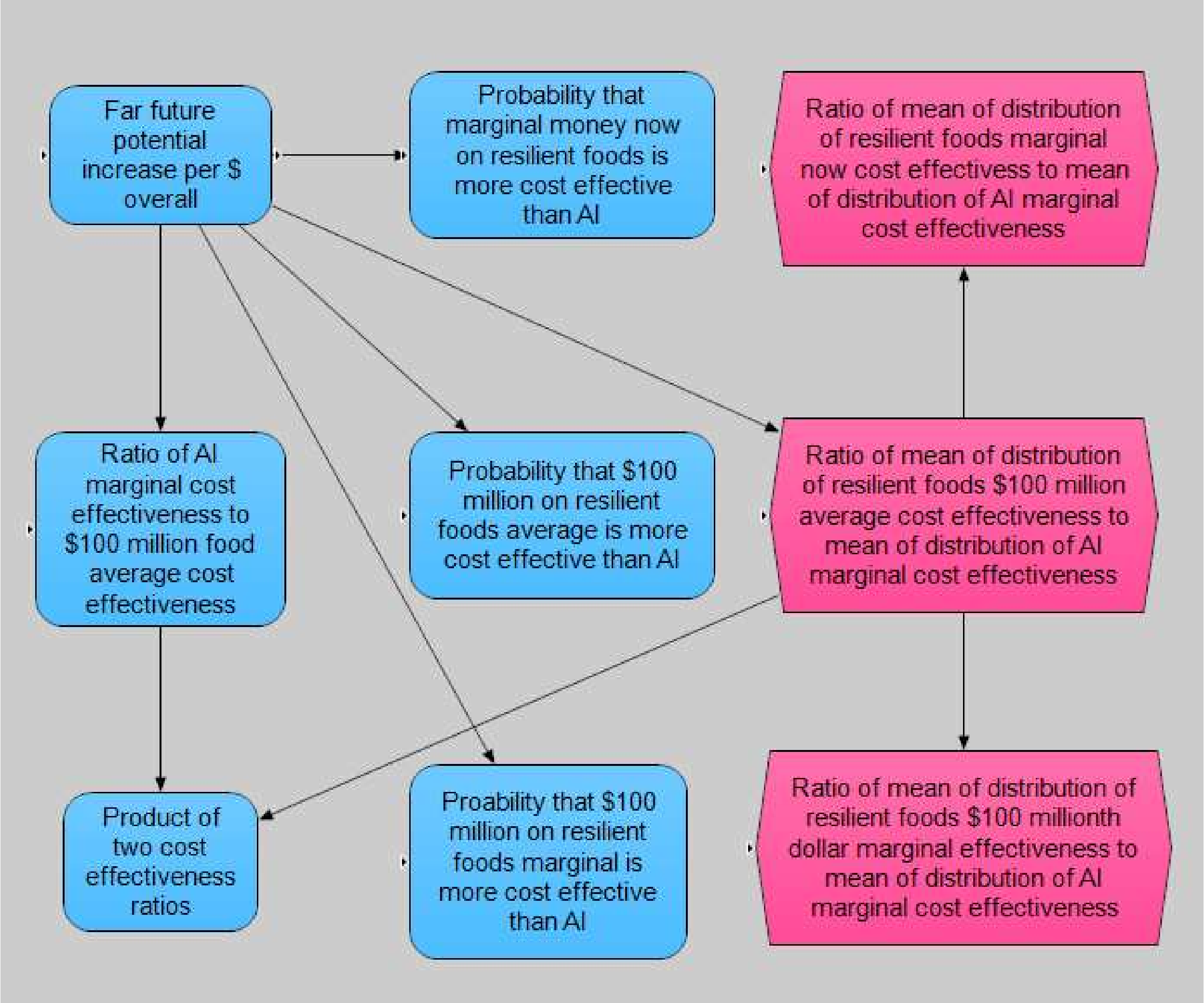

Far future potential increase per $ overall
Probability that marginal money now on resilient foods is more cost effective than AI
Ratio of mean of distribution of resilient foods marginal now cost effectivess to mean of distribution of AI marginal cost effectiveness
Ratio of AI marginal cost effectiveness to $100 million food average cost effectiveness
Probability that $100 million on resilient foods average is more cost effective than AI
Ratio of mean of distribution of resilient foods $100 million average cost effectiveness to mean of distribution of AI marginal cost effectiveness
Product of two cost effectiveness ratios
Proability that $100 million on resilient foods marginal is more cost effective than AI
Ratio of mean of distribution of resilient foods $100 millionth dollar marginal effectiveness to mean of distribution of AI marginal cost effectiveness

Figure 4. S model overall cost effectiveness of resilient foods

2.2 Resilient foods Submodel

Table 1 shows the key input parameters for resilient food for S model and E model. All distributions are lognormal unless otherwise indicated. Though the authors are associated with resilient foods research, two out of four have also published in AGI safety. Furthermore, a diversity of opinions concerning the resilient foods field were incorporated in S model through the utilization of survey data obtained by Alliance to Feed the Earth in Disasters (ALLFED)[9]. The survey solicited responses concerning the likelihood of sun blocking events and likely impact of risk mitigation measures (see Appendix A. for survey text). The value of the long term future is very difficult to quantify, so losses are expressed as a percent.

[9] ALLFED is an NGO that researches technological, financial and preparedness interventions to increase resilience to severe food system shocks. This includes assessment of resilient foods.

Table 1. Resilient food input variables

	S-model		**E-model**	
Input Variable	**5th percentile**	**95th percentile**	**5th percentile**	**95th percentile**
Cost of Planning, R&D for resilient food ($ million)	26	190	10	100
Time horizon of effectiveness of planning and R&D for resilient food (years)	5	50	30	150
Probability per year of a full scale nuclear war between Russia and US [β], [10], [11]	0.006%	2%	0.02%	1.5%
Probability per year of a 10% agricultural shortfall	1%	8%	0.5%	1.5%
Reduction in far future potential due to full scale nuclear war	1%	90%	5%	10%
Reduction in far future potential due to 10% agricultural shortfall	0.03%	40%	0.001%	0.1%
Mitigation of far future impact of full scale nuclear war due to ALLFED so far	0.01%	20%	0.1%	1%
Mitigation of far future impact of 10% agricultural shortfall from ALLFED so far	0.02%	20%	0.1%	1%
Additional mitigation of far future impact of full scale nuclear war due to planning and R&D	3%	90%	10%	50%

[β] Beta distribution used for S model and E model.

[10] S model only considers probability of inadvertent nuclear war.

[11] Values are the average over the index of post 2022 updated nuclear war probabilities.

Additional mitigation of far future impact of 10% agricultural shortfall with planning and R&D	2%	70%	10%	70%
Moral hazard for full scale nuclear war: loss in net effectiveness of interventions	1%	10%	0.1%	1%
Moral hazard for 10% agricultural shortfall: loss in net effectiveness of interventions	0.5%	5%	0.01%	0.1%

The cost of planning, research and development (R&D) for S model is taken from Denkenberger et al. (2016) [35] which was for the 10% food shortfall case ($80 million mean). However, there is a high correlation of preparation for the two catastrophes, so this is assumed to be the cost of the preparation to both scales of catastrophe. E model has lower costs ($40 million mean). The nature of R&D and planning is important in obtaining reasonable cost bounds. Extensive agricultural R&D efforts, such as that performed by the Consultative Group on International Agricultural Research (CGIAR), largely responsible for the 'green revolution', received funding of 7.1 billion USD over 30 years (1971-2001)[62]. However, the advanced nature of the research (for the time), which involved creating novel strains of multiple staple grains, and the breadth of research which presently is split across 15 separate research centers each funded at ~$20 – 50 million USD annually for a total of $580 million USD [63], is much greater than anticipated planning and R&D expected for resilient foods. Resilient foods research & development is likely to focus on several resilient foods already being developed commercially, which implies a more reasonable reference would be the cost to fund one of the 15 research centers (~20 – 50 million per year). Piloting of a resilient food not being commercially developed may be required so another relevant data point is the cost of piloting a chemical process technology. Reported values from several case studies of pilot plants for novel processes indicates costs are commonly in the range of $1 - 20 million USD (average ~10m USD) [64–66]
A final contextual data point is total global agricultural research spending in 2016 which was $47 billion USD, of which 70 countries spent $10 – 100 million [67].

The time horizon of effectiveness of the interventions for S model is also taken from Denkenberger et al (2016) [35] (20 year mean). E model has a significantly longer time horizon corresponding to the rough time constant of the change in our industrial processes, i.e., the rate at which industrial infrastructure is replaced or refurbished (approximate mean of 75 years [68–70]).

The S model uses a beta distribution with parameters ($\boldsymbol{\alpha}$ = 0.61, $\boldsymbol{\beta}$ = 77.19), mean = 0.76%. This distribution is based on Barret et al. 2013 which analyzes only inadvertent full scale nuclear war (attacking when mistakenly thinking your country is being attacked) through a fault tree analysis [7]. The parameters are obtained by tuning a beta prior to the nuclear war probability estimated by Barrett and updating according to zero instances of nuclear war since deployment of the Communication, Command, Control and Intelligence (C3I) system in 1975 (47 years) which inform Barrett 2013 (see Appendix B for details of update). It is important to note that the nuclear war probability only concerns inadvertent nuclear war between Russia and the US and thus ignores other potential avenues to full scale nuclear war such as an intentional first strike or an accidental nuclear explosion [6] and exchange dyads, notably Russia and China or US and China. Inclusion of such factors would significantly increase the probability of "full-scale" nuclear war, but are ignored so as to not overclaim on cost effectiveness. However, we note that twelve out of 16 [71] times that there has been a switch in which is the most militarily powerful country in the world, there has been war (though one should not take that literally for the current situation).

E-model uses a beta distribution with parameters (**α** = 1, **β** = 142) for present day (2022) nuclear war probability. If one assumes a maximum entropy prior and 77 years of no nuclear war (1945 – 2022) with a beta distribution, one gets a mean annual probability of approximately 0.7%. This does not take into account the possibility of China conflicts.

The S model and E model project the probability of inadvertent nuclear war into the future by running the model for an index of discretely updated nuclear war probabilities and taking the mean of the index of runs. The index spans present day (2022) up to the 95th percentile of 'time horizon of effectiveness for resilient foods', 49 years for S model and 149 years for E model, after which it is assumed interventions are no longer effective (Appendix B). Probabilities are then truncated to remove extreme values (Appendix B). The mean probability of nuclear war per year across index runs is ~0.6% for S model and ~0.5% for E model.

Intuitively, one would expect that the probability of 10% shortfalls would be significantly greater than full-scale nuclear war. There are many more potential combinations of regional nuclear war than for full-scale. According to a UK government study [72], extreme weather on multiple continents has ~1% per year chance now and increasing throughout the century. S model mean is 3% per year for 10% of agricultural shortfalls. E model mean is 0.9% per year.

A survey undertaken by ALLFED was sent to 32 GCR researchers and had a 25% response rate (this included responses from two of the authors). The survey questions involved the reduction in far future potential due to several catastrophes, the contribution of ALLFED to resilient foods so far (ALLFED coordinates work on resilient foods), and the additional contribution of spending roughly $100 million to get prepared. The data from the survey was used directly instead of constructing continuous distributions.

The mean estimate of these GCR researchers was 42% reduction in the long-term future of humanity due to full-scale nuclear war if there were no ALLFED, which compares to a 30% [20] estimate by 80,000 Hours, an organization that prioritizes causes. The E model estimate mean was 7%.

The 10% food shortfall catastrophes could result in instability and full-scale nuclear war or other routes to far future impact. The poll of GCR researchers found a mean of 13% reduction in long-term potential of humanity due to these catastrophes. This is similar to the 80,000 Hours' estimate of ~20% [73] for slow extreme climate change. The E model estimate mean was 0.03%. This is much smaller than the other estimates. This would be consistent with the long-term future impact being a significantly greater than linear function of the short-term damage.

The survey also indicated the means of the distributions of percent reduction in far future loss due to ALLFED (and the work done by ALLFED researchers before the organization was officially formed) were 4% for full-scale nuclear war. This value was 0.4% according to E model. Possible mechanisms for impact due to work so far include the people aware of resilient foods already getting the message to decision makers in a catastrophe, decision makers finding the dozen or so papers on resilient foods, or the people in the media who know about resilient foods spreading the message.

ALLFED's contribution to mitigating the 10% agricultural shortfalls was 6% from the survey and 0.4% from E model.

Furthermore, the means of the percent further reduction in far future loss due to full-scale nuclear war due to spending ~$100 million were 30% for the survey and 20% for E model. For the 10% agricultural shortfalls, the mean reductions were 30% for both the survey and E model.

Moral hazard would be if awareness of a food backup plan makes nuclear war more likely or more intense. It is unlikely that, in the heat of the moment, the decision to go to nuclear war (whether accidental, inadvertent, or intentional) would give much consideration to the nontarget countries. However, awareness of a backup plan could result in increased arsenals relative to business as usual, as awareness of the threat of nuclear winter likely contributed to the reduction in arsenals [74]. Mikhail Gorbachev stated that a reason for reducing the nuclear arsenal of the USSR was the studies predicting nuclear winter and therefore destruction outside of the target countries [75]. One can look at how much nuclear arsenals changed while the Cold War was still in effect (after the Cold War, reduced tensions were probably the main reason for reduction in stockpiles). This was ~20% [76]. The perceived consequences of nuclear war changed from hundreds of millions of dead to billions of dead, so roughly an order of magnitude. The reduction in damage from reducing the number of warheads by 20% is significantly lower than 20% because of marginal nuclear weapons targeting lower population and fuel loading density areas. Therefore, the reduction in impact might have been around 10%. Therefore, with an increase in damage with the perception of nuclear winter of approximately 1000% and a reduction in the damage potential due to a smaller arsenal of 10%, the elasticity would be roughly 0.01. Therefore, the moral hazard term of loss in net effectiveness of the interventions would be 1%. S model has a mean moral hazard term of 4%, representing greater elasticity than this calculation due to model uncertainty. E model has a mean moral hazard term of 0.4%, representing lower elasticity than this calculation due to the explanation of reducing arsenals partly due to bankruptcy of the USSR.

For the 10% agricultural shortfalls, S model has a mean 2% loss in net effectiveness, because the moral hazard would apply less strongly to non-nuclear scenarios, such as coincident extreme weather and volcanic eruptions. The corresponding E model value was 0.04%.

2.3 Artificial Intelligence Submodel

The submodel for AGI safety cost-effectiveness was based on work of the Oxford Prioritisation Project, Owen Cotton-Barratt and Daniel Dewey (both while at the Future of Humanity Institute at the University of Oxford). The quantitative Monte Carlo model assesses the impact of additional donations through four steps: 1. Expected value of far future in human -equivalent-wellbeing-adjusted-life-years (HEWALYs), 2. Probability of an AI catastrophe causing human extinction, 3.

Reduction in risk of AI catastrophe per additional researcher, and 4. Cost of additional researcher. Although unpublished, significant effort was made to assess model and theoretical challenges.[12] Furthermore due to the small size of the existential risk research community and the even smaller size with the relevant knowledge and skill to perform such cost effectiveness assessments, this was the only credible and publicly available model assessing the impact of AGI safety on the far future. Distributions are lognormal except where otherwise indicated.

The first step is estimating the reduction in the value of the long-term future due to AGI risk bearing in mind all the work that is expected to be done to reduce that risk. This allows a marginal cost-effectiveness analysis. Unlike in the case of agricultural catastrophes, if there is an AGI catastrophe,[13] it is likely to cause human extinction according to [78] or result in the loss of irrecoverable loss of humanity's potential [18]. One existential scenario could involve AGI developing atomically precise manufacturing and remaking the surface of the earth (and potentially the galaxy and beyond) into whatever its objective function specifies [78]. However, there are other possible failure modes [51].

Method 1 first involves estimating the total existential risk (over the next 100 years). This time horizon is significantly longer than that considered for resilient foods, partly because the utility of resilient foods investments would be cut short given the appearance of AGI. This is due to either AGI destroying humanity or enabling far greater technology to withstand agricultural catastrophes. Also, the investment in AGI safety is seen more as an entire career, which would be relevant further into the future. But overall, this is likely pessimistic towards the relative cost effectiveness of resilient foods. Total existential risk is described by a beta distribution with parameters ($\boldsymbol{\alpha}$ = 1.5, $\boldsymbol{\beta}$ = 8). The upper bound is based on Martin Rees [82] (50%). Other sources, including a Metaculus forecasting poll[14] with 740 predictions argue for much smaller percentages[83,84]. The resulting mean is 16% (increased from the original model) (Table 2), which agrees with the total existential risk estimated by Toby Ord [44]. Next there is the percentage of total existential risk associated with developing AGI, which is based on many existential risk researchers believing that AGI is the dominant existential risk (mean of 26%), a beta distribution with parameters ($\boldsymbol{\alpha}$ = 2, $\boldsymbol{\beta}$ = 5) is used (Table 2). They are multiplied together. Method 2 is estimating the reduction in long-term value of humanity from AGI directly (mean of 8%) (increased from the original model) and also described by a beta distribution, parameters ($\boldsymbol{\alpha}$ = 1.2, $\boldsymbol{\beta}$ = 15) [51] (Table 2). Two thirds weight

[12] Discussion of limitations, and assumptions of the Machine Intelligence Research Institute cost effectiveness model available at Oxford Prioritisation Project website [57].

[13] AGI Catastrophe is used to describe scenarios in which AGI permanently and drastically curtails the potential of humanity. There exists a range of views on the plausibility and mechanism of such catastrophic events. Common arguments include: strong optimizers pursuing non human aligned goals to devastating consequences e.g. turning all matter into paper clips [77,78]; incorrect goal measurement choices, i.e. substituting easy to measure goals vs hard to measure goals e.g. reducing reported crime vs actually preventing crime, which allow for slower catastrophes to unfold as compounding of incorrectly measured goals lead to catastrophic outcomes [79]. Super intelligence may not be required to pose catastrophic risk, 'narrow' AI may possess the necessary capabilities [80], and even if alignment is successful, AGI fueled increases in the speed of development of new technologies could lead to vastly more powerful or dangerous technologies being created and utilized by humans before improvements in coordination and rationality have occurred, feeding into the fragile world theory [81].

[14] Metaculus is an online prediction platform that aggregates forecasts on real world events from a community of forecasters.

is assigned to method 1, and one third to method 2, since method 1 forces more thought by breaking it down into two steps.

Table 2. AGI safety input variables

Input Variable	**5th percentile**	**95th percentile**
Method 1: Total existential risk over the next 100 years[β]	2.3%	38%
Method 1: Percentage of total existential risk associated with developing highly capable AI systems[β]	6.4%	58%
Method 2: Total existential risk associated with developing highly capable AI systems[β]	0.8%	20%
Size of the research community working on safety by the time we develop those potentially risky AGI systems[η]	90	1000
Effect of adding a researcher to the community now (and for their career), in terms of the total number of researchers that will be added to the eventual community	0.75	5
If double the total amount of work that would be done on AGI safety, what additional percentage of the bad scenarios should we expect to avert?	0.1%	10%
Cost of a researcher to enter field of AGI safety for the rest of their career	3,000,000	10,000,000

The next step is estimating how much an additional AGI safety researcher reduces risk, which is incorporated into the model by utilizing relevant components from the model developed by Owen Cotton-Barratt and Daniel Dewey.

First, it is estimated that the productivity of one AGI safety researcher, in terms of the fraction of AGI risk they reduce. To do this, a value is assigned to the size of the research community working on safety by the time those potentially risky AGI systems are developed, mean 500 (normal distribution) (Table 2). Then, a value is assigned to the following: if the total amount of work that would be done on AGI safety were doubled, what additional fraction of the bad scenarios would be averted? This has a mean of 3% (Table 2). Then the average productivity of each of the 500 researchers added is: 3%/500. Assuming returns are logarithmic [85], the marginal value of the first of 500 researchers would be approximately 1.4 times as much as the average (increased from the original model). This does do not account for possible changes in researcher quality. This is the direct effect of adding a researcher today.

[β] Beta Distribution

[η] Normal Distribution

However, there is an indirect effect: adding a researcher today may eventually cause more or fewer people to eventually be in this area. A value is assigned to that spillover effect: how many extra researchers eventually enter the field if one is added today (this equals one if there is no spillover effect, and is greater/less than 1 if the effect is positive/negative). The mean is 2.3 (Table 2). The direct effect is multiplied by this number above to get the net effect of an additional researcher today. Of course, these additional researchers would need to be paid, so one could argue that this is optimistic from the perspective of AGI safety cost effectiveness.

The last step is how much an additional researcher costs: mean of 6 million USDs as the cost over his/her career (increased from the original model).

3 RESULTS & DISCUSSION

3.1 Results

In order to convert average cost effectiveness to marginal for resilient foods, again logarithmic returns are assumed, which results in the marginal cost effectiveness being just one divided by the cumulative money spent. In this case, the marginal cost effectiveness on the last dollar for resilient foods would be about one fifth the average cost-effectiveness for S model and one fourth for E model. For funding resilient foods at the margin right now, an estimate is needed of the cumulative money spent on resilient foods. Under $1 million equivalent (mostly volunteer time) has been spent so far directly on this effort, nearly all by the Alliance to Feed the Earth in Disasters (ALLFED)[86]. Therefore, cost-effectiveness of the marginal dollar now is about 5 times greater than average of $100 million assuming logarithmic returns for S model, and 3 times greater than average of $40 million for E model. Ratios of mean cost effectivenesses are reported in Table 3.[15]

Table 3. shows the ranges of the far future potential increase per $ due to resilient foods average over ~$100 million and the far future potential increase per $ due to AGI safety research at the $3 billion margin[16]. Figure 5 shows the cost-effectiveness distribution for AGI safety, S model and E model for resilient foods. The ratios of the 95th and 5th percentiles for each model indicate the variance of the model output. The index average ratios are 1000 for S model, 30 for E model, and 500 AGI safety. Because the variance of S model is high, the mean cost-effectiveness is high, partly driven by the small chance of very high cost-effectiveness.

Table 3. Comparison of cost effectiveness of resilient foods (S-model and E model) and AGI safety research

[15] Of course a very large amount of money has been spent on trying to prevent nuclear war. More relevant, money has been spent developing resilient foods for other reasons, such as mushrooms and natural gas digesting bacteria. This could easily be tens of millions of dollars that would have needed to be spent for catastrophe preparation. So this would be relevant for the marginal $100 million. However, there are very high value interventions that should be done first, such as preparing to exploit mass/social media in a catastrophe to get the right people to know about resilient foods. Though the resilient foods would not work as well as with $100 million of R&D, simply having the leaders of countries know about them and implement them in their own countries without trade could still significantly increase the chance of retaining civilization. The cost of these first interventions would be very low, so it would be very high cost effectiveness.

[16] Over $1 billion has already been pledged to AGI safety research, making $3 billion appear conservative [87–89].

Output	**5th percentile**	**95th percentile**
Far future potential increase per $ due to resilient foods average over ~$100 million S model	2x10^{-12}	2x10^{-9}
Far future potential increase per $ due to resilient foods average over ~$100 million E model	3x10^{-11}	8x10^{-10}
Far future potential increase per $ AGI safety research at the $3 billion margin (same for both models)	2x10^{-14}	1x10^{-11}

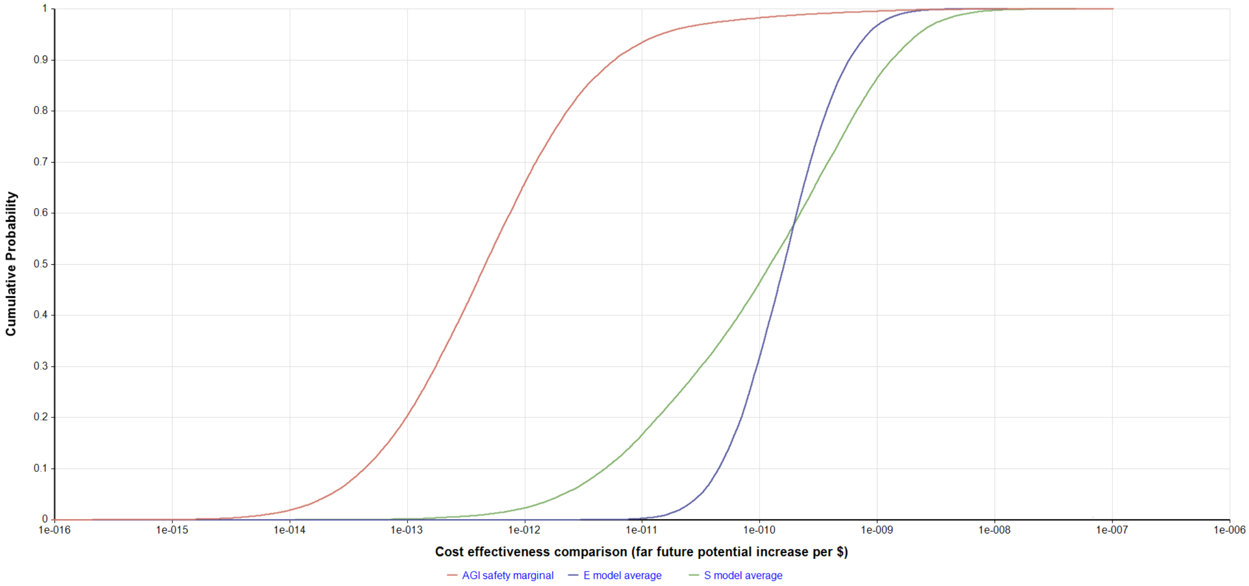


Figure 5. Far future potential increase per $ overall average over ~$100 million S model, overall average over ~$40 million E model, and AGI safety research at the $3 billion margin. More cost effective is further to the right.

Comparing to AGI safety at the margin, S model yields the 100 millionth dollar on resilient foods being 6 times more cost effective, the average $100 million on resilient foods being 20 times more cost effective, and the marginal dollar now on resilient foods being 400 times more cost effective (see Table 4). E model yields the 100 millionth dollar on resilient foods being 3 times more cost effective, the average $100 million on resilient foods being 10 times more cost effective, and the marginal dollar now on resilient foods being 100 times more cost effective. One way of thinking about the high marginal cost effectiveness now is spending some money to figure out if more money is justified: value of information [90]. Given the sensitivity of the ratios of the means to the variance, more robust is likely the probabilities that one is more cost effective than the other. Comparing to AGI safety at the margin, S model finds ~84% probability that the 100 millionth dollar on resilient foods is more cost effective, ~92% probability that the average $100 million on resilient foods is more cost effective, and ~98% probability that the marginal dollar now on resilient foods is more cost effective (see Table 4). E model finds ~93% probability that the 100 millionth dollar on resilient foods is more cost effective, ~97% probability that the average $100 million on resilient foods is more cost effective, and ~99% probability that the marginal dollar now on resilient foods is more cost effective.

Table 4. Ratios of resilient foods cost effectiveness to AGI safety and confidence values.

	S-model		**E-model**	
Scenario	**Ratio of resilient foods mean cost effectiveness to AGI safety mean cost effectiveness**	**Confidence that resilient foods is more cost effective than AGI safety**	**Ratio of resilient foods mean cost effectiveness to AGI safety mean cost effectiveness**	**Confidence that resilient foods is more cost effective than AGI safety**
100 millionth dollar to resilient foods	6	84%	4	93%
$100 million average to resilient foods	20	92%	10	97%
Money to resilient foods at the margin now	400	98%	100	99%

Overall, the mean cost-effectivenesses of the two resilient foods models were similar. Because of the smaller variance in the E model distributions, there was greater confidence that resilient foods are more cost-effective than AGI safety. Another large difference is that S model found that 10% agricultural shortfalls are similar cost effectiveness for the far future as full scale nuclear war. This was because the greater probability of these catastrophes counteracted the smaller far future impact. However, E model rated the cost-effectiveness of the 10% shortfalls as two orders of magnitude lower than for full-scale nuclear war.

Being prepared for agricultural catastrophes might protect against unknown risks, meaning the cost-effectiveness would increase. According to the average across indexed nuclear war probabilities in the S model, every year acceleration in preparation for resilient foods would increase the long-term value of humanity by 0.001% to 0.6% (mean of 0.2%). The corresponding E model numbers are 0.0038% to 0.02% (mean of 0.008%). Either way, there is great urgency to prepare.

It is not required for resilient foods to be more cost effective than AGI safety in order to fund resilient foods on a large scale. Funding of the existential risk community goes to other causes, notably avoiding an engineered pandemic. One estimate of cost effectiveness of biosecurity was significantly lower than for AGI safety and resilient foods, but the authors were being very conservative [55]. Another area of existential risk that has received significant investment is asteroid impact [53]. Again, the cost-effectiveness is much lower than for resilient foods.

Supposing a defense in depth framework[17] [91] the case for resilient foods is even stronger compared to other sun blocking catastrophe interventions. Resilient foods mainly function on the resilience level, i.e., increase the ability to prevent a sun blocking catastrophe from eliminating all of humanity or permanently curtailing humanity's potential, which is comparatively underfunded compared to prevention and response levels. This is especially the case for nuclear security where the majority of funding is provided at the prevention level through initiatives such as the International Atomic Energy Agency International non-proliferation treaty, national level nuclear security activities and other diplomacy-based risk reduction strategies.[18]

The importance, tractability, neglectedness (ITN) framework [95] is useful for screening cause areas. Importance is the potential impact of the risk on the long-term future. Neglectedness quantifies how much effort is being put into reducing the risk. Unfortunately, this framework cannot be directly applied to interventions. This is because addressing a risk could have many potential interventions. Still, some semi quantitative insights can be gleaned. The importance of AGI is larger than agricultural catastrophes, but resilient foods is significantly more neglected. Resilient foods appear to have unusually low-hanging fruit still available, which without additional funding would not likely be taken by existing actors such as Single Celled Protein (SCP) companies (Unibio, Calysta, Solar Foods, Novonutrients, etc.)[27] and cellulosic sugar companies (Comet Bio, Renmatix, etc.)[26] or governmental disaster planning agencies (FEMA, UNDRR etc.) as they are not focusing on GCRs. This is not the case for AI safety. Concretely, marginal funds for resilient foods could go into highly leveraged activities such as developing rapid scale-up and infrastructure retrofit[19] methodologies or GCR response planning, which, without additional funding might not happen, or only happen significantly later. There is no analog in AI

[17] The defense in depth existential risk framework describes risks according to steps a catastrophe must progress through to scale to an existential risk and corresponding potential intervention points. The three layers are prevention, response and resilience. This framework implies that it is usually most cost effective to fund all three layers equally in order to reduce existential risk due to higher marginal cost effectiveness opportunities existing at each level and because catastrophes must progress through all three levels to become an existential threat. As such layering the most cost efficient interventions from all three levels (prevention, response and resilience) will lead to the greatest decrease in total probability of a catastrophe being able to progress through all three levels to extinction per dollar.[91]

[18] The National Nuclear Science Administration will receive USD $ 2 billion in 2021 for nonproliferation activities.[92], open philanthropy's report [93] mainly describes prevention level funding and GCRI nuclear security research [94] predominantly looks at prevention level interventions e.g. winter safe deterrents.

[19] A specific example would be the development of plans for retrofitting SCP facilities to produce human food (instead of fish feed) during disasters. This would be highly leveraged because it could utilize commercially viable infrastructure that is being developed to service protein requirements of rapidly growing aquaculture markets [96–98]

safety that would not be taken by the billions of dollars committed funding by default. Tractability is more difficult to ascertain, but if it were similar, this would predict similar cost-effectivenesses.

The AGI safety submodel can also be used to estimate the cost effectiveness of saving lives in the present generation. With the assumption that an existential catastrophe with AI means the loss of 9 billion humans, the cost effectiveness of AGI safety now is $16-$12,000 per expected life saved. This is not nearly as cost-effective as resilient foods ($0.20-$400 for only 10% global food production shortfalls) [35], but it is generally lower cost than GiveWell estimates for global health interventions: $900-$7,000 [99]. Technically there would be a cost associated with producing the resilient foods in a catastrophe (though mosquito bed nets are benefiting from past costs associated with pesticide development, etc.). However, the expenditure on resilient foods would actually be less than the expenditure on stored food alone because the price of stored food would go so high, even though stored food would not be able to feed as many people [18]. Therefore, one could argue that the cost of saving a life with resilient foods is actually negative.

Since AGI safety appears to be underfunded from the present generation perspective, it would be extremely underfunded when taking into account future generations. If this were corrected with more funding, then in order to have similar cost-effectiveness with resilient foods, more funding for resilient foods would be justified. Indeed, in order to fund resilient foods just from a current generation perspective at a level of similar cost-effectiveness to global poverty interventions, billions of dollars of resilient foods funding would be justified [35]. Much more funding would be justified if valuing future generations.

3.2 Funding Considerations

If one agrees that resilient foods should be in the existential risk portfolio, the next question is how to allocate funding to the different causes over time. For AGI safety, there are arguments both for funding now and funding later [100]. For resilient foods, since most of the catastrophes could happen right away and relevant knowledge may be more reliable to attain, there is significantly greater urgency to fund resilient foods now. Furthermore, it is relatively more effective to scale up the funding quickly because, through requests for proposals, the effort could co-opt relevant expertise that already exists (e.g. in the different foods, such as biofuel experts who know how to turn fiber into sugar). Since the value of the far future has not been monetized, traditional cost-effectiveness metrics cannot be used such as the benefit to cost ratio, net present value, payback time, and return on investment. However, in the case of saving expected lives in the present generation for the global case and 10% food shortfalls, one study indicated return on investment was from 100% to 5,000,000% per year [35] based on monetized life savings. This suggests that the $100 million or so for resilient foods should be mostly spent in the next few years to optimally reduce existential risk (a smaller amount would maintain preparedness in the future). Since AGI safety funding is now about $10 million per year [101], this would indicate more funding for resilient foods than AGI safety in the near term.

Another way of thinking about the timing of money going into resilient foods versus AGI safety is whether investments could actually be monetized. One way of doing this is offering insurance policies for catastrophes that could be revenue-generating. Another possible way is funding preparedness and making an agreement with the government. Developed country governments would likely pay exorbitant amounts to feed their citizens in a catastrophe with only stored food [36]. The agreement could be that the donors would be paid a proportion of savings. Then in expectation, much more money could be put into AGI safety later.

Research and intervention funding is influenced by political [102] and social factors [103][104][105] and prone to various biases, most notably temporal discounting and scope insensitivity [106]. Consequently, allocation of funding for interventions to low probability, high impact events are prone to temporal discounting and scope insensitivity effects, resulting in funding being insufficient to the risk posed. The methodology presented aims to address this by providing a measure that adequately considers the scope and timeframe of the risk and resulting cost effectiveness of interventions, improving the rigor of risk mitigation funding efforts. Used in tandem with short term-oriented analysis, the model presented may also help identify high leverage opportunities which have meaningful impact on short term and long term time frames. Preparation for 10% agricultural short fall events such as multiple breadbasket failure or VEI Seven volcano events (which would result in partial sun obscuring) provide good examples, being likely to occur this century and cause large impact [107]. Preparation for these events also provides an incremental improvement in preparation for 100% agricultural collapse scenarios. Another example could be the development of certain resilient foods (e.g. SCP) to help meet the massively increasing protein demand within environmental limitations, while also helping add resilience to food production that would be valuable in sun obscuring catastrophes [108].

Funding for resilient foods could be justified by only considering the cost effectiveness of 10% agricultural loss scenarios (3% - 30% loss of agricultural output). The high likelihood of such events occurring in the short term and the significant impacts, require no belief in far future scenarios and implies funding outside of the normal existential risk sources could be justified. For instance humanitarian aid, which encapsulates poverty reduction through education, economic development, infrastructure development and improving agriculture [109] receives ~ 30 billion USD annually [110]. The impact of 10% food system shocks i.e. food price spikes and cascading impacts such as political instability, increased conflict and economic impacts fall disproportionately on the global poor, negatively impacting humanitarian aid work [43,72]. Consequently funding resilient foods interventions via humanitarian aid funding would prevent acute shocks from worsening conditions for the global poor, saving lives cost effectively ($0.20-$400 per expected life saved) [35].

3.3 Theory and Model Uncertainty

The complex sociotechnical systems investigated in this model introduce a variety of model and theory uncertainties [111]. These uncertainties exist outside of the model. Theoretical uncertainty represents the dominant form of uncertainty having a greater impact than model uncertainty. Below the theory and model uncertainties are explored and the corresponding limitations they put on the results of the model.

3.3.1 *Theory uncertainty*

Predicting the impact of any intervention on the far future contains inherent epistemic uncertainty that is not readily resolved. Cluelessness, the inability to predict the vast majority of consequences brought about by undertaking a given action, presents an epistemic hurdle to determining to what extent we can meaningfully change the future. Agreeing that 'Simple' cluelessness[20] is illusory if one accepts the notion of subjective criterion of consequence-betterness of an action, the potential sources of 'complex' cluelessness are accepted[21] apply to the sociotechnical systems discussed and that this introduces a level of theoretical uncertainty outside of the model [112].

One pertinent aspect of theory uncertainty is whether the proposed interventions could result in a net negative impact. Overall, net negative impacts are unlikely as much previous work on existential risk mitigation indicates [44,55,113–115]. Furthermore, negative impacts would be possible for both resilient foods and AGI safety and there is no obvious reason why either would be more affected. Therefore, for the purpose of comparing cost-effectiveness, it is a defensible simplification to focus on cases of net-positive impacts. Though beyond the scope of this paper, a more in-depth treatment of this epistemic issue would make a valuable addition to the literature. In this vein, claims made about the long-term future are even more uncertain. Time frames of billions of years make it likely for unknown extremely negative (or positive) events[22] which would outweigh the value created by any deliberate action to occur. To overcome such extremely high impact events and make long termist arguments plausible requires a credence in unlikely but extremely valuable outcomes for given actions [116]. In summation, there are limitations to the certainty one can have on the impact of the interventions laid out in this paper on the long-term future, and as such results should be interpreted in an epistemically reserved manner. A potentially fruitful future research direction would be to model the cost effectiveness of resilient foods over a shorter term (< 1000 years), in order to overcome some of the epistemic uncertainty related to determining the value of the long term future and strengthen the case for resilient foods as a portion of the Global Catastrophic Risk portfolio.

3.3.2 *Model uncertainty*

Survey structure, solicitation method and field specific epistemic challenges are likely to bias survey responses (Survey available as Appendix A., specific responses available in Analytica model.). The biases present in the survey and corresponding impact on results are discussed below.

The low survey turn out (8 out of 32) implies a potential response bias. Demand characteristic bias [117] is likely present as the number of respondents was small and respondents volunteered and could have deduced for the purpose of the survey. Consequently social conformation could motivate favorable responses towards the deduced hypothesis, this would lead to responses

[20] Simple' cluelessness argues that it is impossible to objectively predict the impact of any action due to compounding events resulting in unforeseeable effects responsible for the majority of the impact of said action, these effects are of unknown sign i.e. can be either positive or negative [112].

[21] 'Complex' cluelessness relates to the epistemic challenge of determining the credence one can have in the outcome of any given action which is likely to contain a variety of knock on systemic effects of unknown valence that could compound to create a larger effect than the initial action [112]

[22] Described in Tarsney's paper *The Epistemic Challenge to Longtermism,* extremely negative (or positive) events introduce greater uncertainty in our ability to impact the long run future. Extremely negative (or positive) events refer to unknown but extremely impactful events that occur at extremely low frequencies, but given enough time >1,000,000 years occur at a frequency large enough to make the utility of such an event outstrip any potential impact a given action could have on the far future. In order to overcome such events and retain some level of agency over the future one must believe in the equivalently unlikely but extremely impactful outcomes some actions can have [116]

favorable to resilient foods e.g. increase value of the far future due to resilient foods, or lower ratings of moral hazard resilient foods.

The accuracy of GCR expert respondents is difficult ascertain, thus judgements should be treated high uncertainty. The survey questions require an understanding of complex sociotechnical systems, spanning multiple disciplines. It is difficult to ascertain whether respondents' knowledge of these varied fields was commensurate with faithfully answering questions and not substituting difficult questions for simplified questions [118]. Determining the accuracy of GCR experts according to the Cochrane-Weiss-Shanteau index [119], would provide a potential method to counteract challenges in obtaining reliable assessments of GCR and would be valuable to informing future GCR priorities research. Inclusion of responses from four affiliates of the organization that conducted the survey (ALLFED) has the potential to skew results: inside view biases might lead to overestimation of impact of mitigation measures, or the estimation of how rapidly this preparation could occur in a given amount of time and confirmation bias may also see members overestimating the efficacy of resilient foods interventions or the reduction in far future potential from nuclear war. In general, responses from ALLFED members would be expected to underestimate uncertainty and overestimate negative impacts of nuclear war on humanity's far future and the impact of resilient foods on mitigating such impacts. Incorporation of survey results from two of the authors is also a potential source of bias.

Question framing could also have introduced bias, e.g. "estimate of the reduction in potential of humanity" assumes net-negative effects from nuclear war, while "percent mitigation" assumes some mitigation would exist. These biases would skew results of the model towards increased cost effectiveness of resilient foods.
Although there exist biases within the survey data utilized in S model, the inclusion of said data provides a more diverse representation of perspectives than solely the authors creating distributions. The inclusion of more sources of data allows the balancing of the various biases each of the respondents has, smoothing out the overall effect of biases.

Lognormal distributions were generally chosen for values that could not be negative, spanned large ranges and did not have an upper limit, because these are relatively common in physical systems. [120,121]. Lognormals are well suited for describing multiplicative processes such as products of weakly- or uncorrelated probabilities, [122] representing the multiplicative counterpart of the Gaussian distribution in the central limit theorem. While heavy-tailed it is not as extreme as a power-law, corresponding to estimates where there are at least some rough consensus of magnitude.

An earlier version of this model focused only on the probability of collapse of civilization and the probability that it is not recovered. However, it was then realized that there were other routes to impact on the long-term future. Another possible way of framing impact for resilient foods would be to estimate the effect of spending money ahead of time on accelerating when large-scale resilient food production would be available after the catastrophe.

It is not clear how one should proceed to update the probability of nuclear war into the future given the uncertainty of the impacts of inadvertent full scale nuclear war. Naively, one might assume that if nuclear war occurs, the likelihood of successive exchanges would increase. While this makes sense from a Bayesian perspective, the physical reality of this is not clear. The initial exchange may use and destroy (through counterforce targeting) the vast majority of warheads making follow-up exchanges of a similar magnitude no longer possible for some time. Alternatively, impacts of the exchange may degrade civilization so much that the capacity to initiate nuclear exchange in the future is massively reduced. A rigorous treatment of this source of uncertainty would make valuable research; however, is considered beyond the scope of the paper. The presented model avoids this issue by assuming that the status quo of no inadvertent nuclear war is continued, which enforces a reduction in the probability of nuclear war with each successive year. Thus, the average over the indexes for both models are an underestimate, representing a lower bound of the possible probability of inadvertent nuclear war between Russia and the US obtainable from the proposed prior distributions. When considered with the favorable ratio of cost effectiveness against AGI safety (Table 4.) this further indicates the potential value of resilient foods.

Utilizing different model structures likely confounded biases between the resilient foods and AGI safety submodels. For instance, the use of different far future impact per unit metrics e.g. far-future impact of additional dollars spent for the resilient foods submodel, compared to far future impact per additional researchers means that biases are inconsistent between models and thus cannot be cancelled out as they would be if model structures were identical. It is still valid to compare the two as it is possible to convert cost per researcher to a dollar value i.e. salary per researcher plus overhead, but greater uncertainty is expected due to the inequivalent biases present in the two submodels of the model.

In the model developed here, AGI and resilient foods submodels are considered *independent* of one another. This can be thought of as investigating two separate universes in which either a sunblocking, or an AGI catastrophe occurs. In reality, the catastrophes that the submodels are investigating occur in the same universe, as such one should expect the impacts of one catastrophe to influence the likelihood and impacts of the other. For example, if a nuclear war occurs before AGI is created this could result in a civilization collapse that prevents AGI from ever being developed, this would majorly influence future impacts of AGI (positive or negative). As such, the two submodels should be *dependent.* The complex nature of such interactions makes it difficult to rigorously model the required dependencies further highlighting model uncertainty. Exploring alternate model structures and associated theory such that the dependencies of submodels are considered would be valuable future work.

Another highly relevant model to be explored would be an agent-based model. Agent-based models would be well suited to investigate the complex sociotechnical systems described and capture emergent behavior which could occur, such a model could be described at region, state or country level. Furthermore, layering of this model and following models investigating the cost effectiveness of resilient foods compared to AGI safety improve the robustness of results by minimizing the impacts of model uncertainty for each model.

3.3.3 Uncertainty Summary

If one wishes to prioritize actions according to far future impacts by using methodologies similar to the model presented, significant theoretical uncertainty must be accepted. Consequently, the reasonableness of the outputs of the model should be assessed according to the defensibility of the structure and choice of probability distributions. The model structure and probability distributions chosen are defensible, being consistent with literature sources and suitable real world examples. Readers with significantly different priors are encouraged to use the model to explore different probability distributions and resulting model outputs. The intention of the model is to provide a useful starting point for investigating the long term case of resilient foods and its inclusion in the existential risk reduction portfolio in order to catalyze future work in the area.

3.4 Sensitivity Analysis

Parameter uncertainty, i.e. uncertainty captured within the model, was investigated through the use of the Analytica importance analysis function, which identifies input variables with uncertainties that most affect model outputs. This analysis uses the absolute rank-order correlation between each input and the output as an indication of the strength of monotonic relations between each uncertain input and a selected output, both linear and otherwise [123,124].

Parameter sensitivity of S model and E model was investigated using the Analytica importance analysis function. Analysis was focused on the resilient foods submodel i.e. full-scale nuclear war and 10% agricultural loss. Parameter sensitivity within the Artificial Intelligence Submodel was not investigated as the submodel was adapted from previous work by the Oxford Prioritisation Project, which considered uncertainties within the AGI safety cost effectiveness submodel [57].

Key output nodes summarized in Table 4 were not amenable to investigation directly using the importance analysis function due to the node outputs being non probabilistic, a result of calculating the ratio of means (the Analytica importance analysis function requires the variable be a probabilistic variable to perform absolute rank-order correlation). Thus, the previous node in the models far future potential increase per $ due to resilient foods was used to investigate the importance of input variables of the resilient foods submodel.

Importance analysis of node: far future potential increase per $ due to resilient foods showed S model had greatest sensitivity to the input variable Reduction in far future potential due to 10% agricultural shortfall (Figure 6). E model showed greatest sensitivity to input variable Probability per year of full-scale nuclear war (Figure 7).

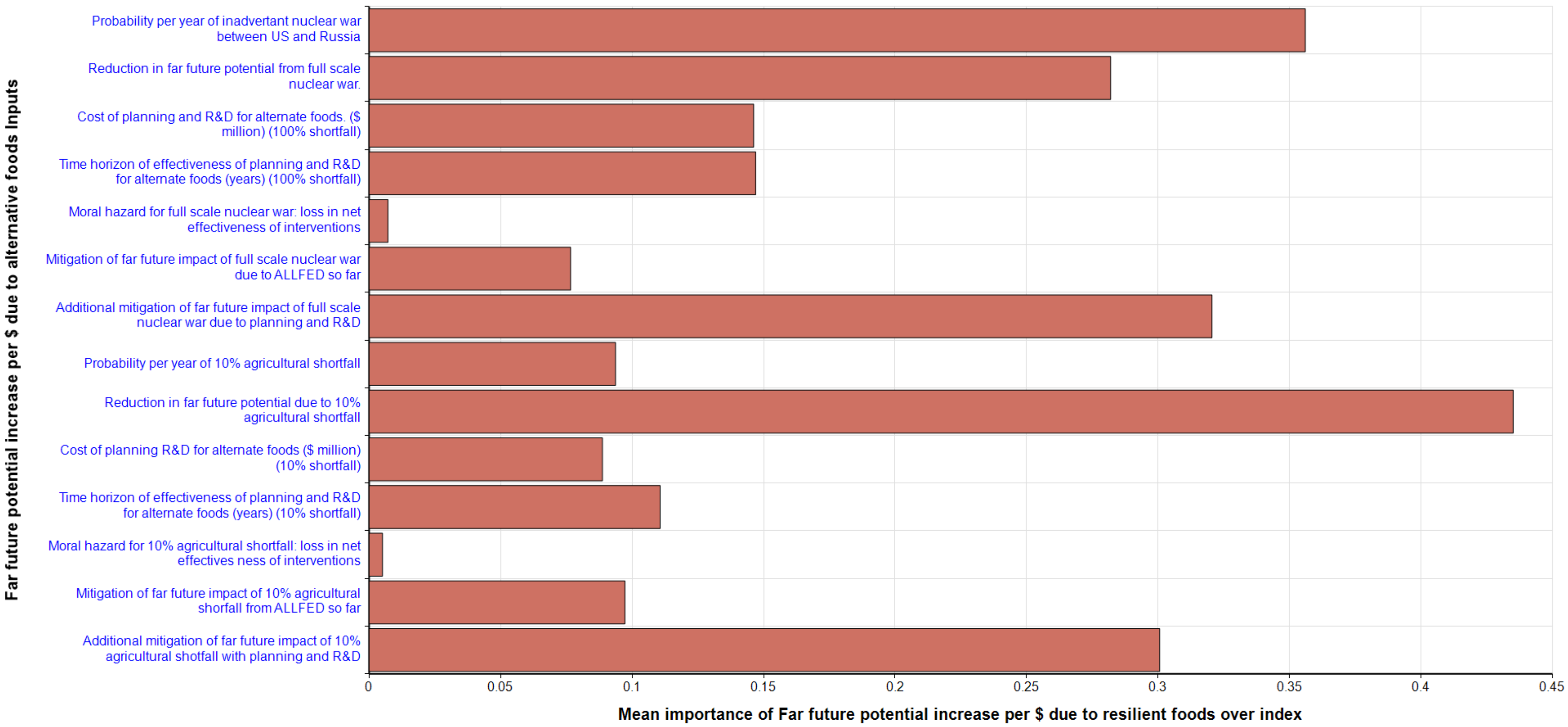


Figure 6. Mean importance analysis results for far future potential increase per $ due to resilient foods over index for S model. Importance is the absolute rank order correlation between an uncertain input and selected output; it is used to measure the degree of association between two variables. An importance of 0.43 means 43% of samples show an association between the input and output.

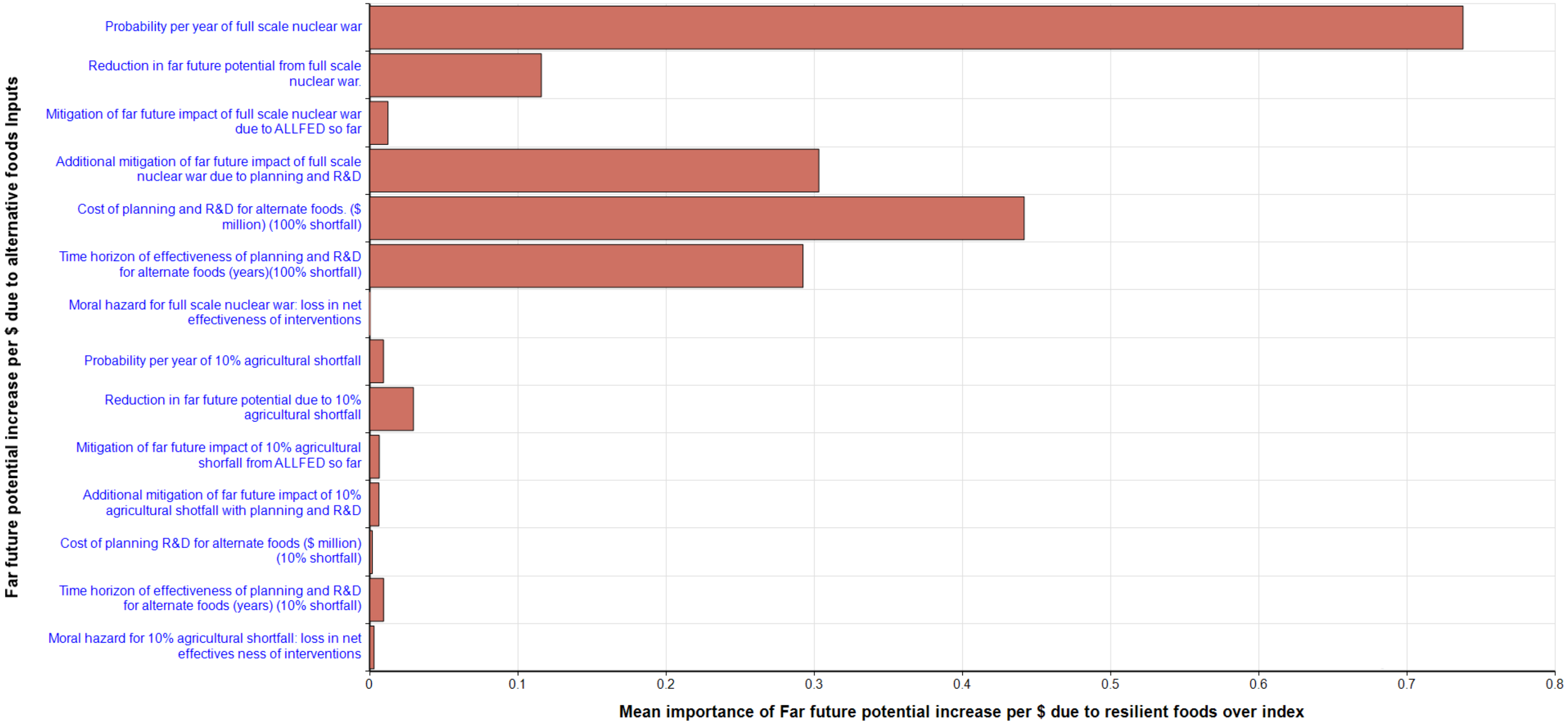


Figure 7. Mean importance analysis results for far future potential increase per $ due to resilient foods over index for E model. Importance is the absolute rank order correlation between an uncertain input and selected output; it is used to measure the degree of association between two variables. An importance of 0.74 means 74% of samples show an association between the input and output.

Successive rounds of parametric analysis were performed to determine combinations of input parameters sufficiently unfavorable to resilient foods, until key cost effectiveness ratios (Table 3) switched to favoring AGI safety for both models. Unfavorable input values were limited to 5th or 95th percentile values of original input distributions. For probability per year of full scale nuclear war variable the most unfavourable value in the index of probabilities was selected. S model required three unfavorable input parameters to switch to AGI safety being more cost effective than resilient foods at the margin now while E model required five unfavorable input variables (see Table 5). This is partly because E model generally has smaller variation in input variables, so more variables have to be changed in order to reverse the results of which cause is most cost-effective.

Table 5. Combination of input variables resulting in AGI safety being more cost effective than resilient foods at the margin now.

Input Variable	**S-model**	**E-model**
Cost of Planning, R&D for resilient foods ($ million)	-	100
Time horizon of effectiveness of planning and R&D for resilient foods (years)	-	30
Probability per year of a full scale nuclear war	0.0048%	0.018%
Reduction in far future potential due to 10% agricultural shortfall	5.0x10^{-5}	1.0x10^{5}
Additional mitigation of far future impact of full scale nuclear war due to planning and R&D	0.030	0.10
Output variable		
Ratio of money to resilient foods at the margin now mean cost effectiveness to AGI safety mean cost effectiveness	**0.22**	**0.44**

A robustness analysis for the confidence values of one type of intervention being more cost-effective than the other was not performed. This would be more insensitive to the variance in the distributions than the ratios of the mean cost-effectivenesses. Since the variance in S model is so broad, it would require fewer variables to be made pessimistic in order for it to be less than 50% confident that it is more cost-effective than AGI safety (than the number of variables required for the ratio of the means in Table 4). Conversely, since the variance in E model is narrower than for

AGI safety, it would require fewer variables to be made pessimistic in order for it to be less than 50% confident that it is more cost-effective than AGI safety.

One future work project would be to analyze the cost-effectiveness of AGI safety and resilient foods in terms of species saved. Unaligned AGI could cause the extinction of nearly all life on earth. If there were loss of much sunlight, this would likely cause some extinctions directly, but also if there were mass human starvation, humans would likely eat many species to extinction. Therefore, being able to meet human needs would save species (and little additional food would be required to keep many species alive that would have gone extinct from the direct catastrophe). These cost effectivenesses could be compared to the cost effectiveness of conventional methods of saving species such as preserving habitat.

Research for the actual preparedness should be done, including better quantifying the scale up speed of resilient foods and the associated cost. Since there would be some sunlight available in most of these scenarios, seaweed and greenhouses could be promising food sources [32]. This should be done under a number of cooperation scenarios, such as global cooperation, cooperation only within countries, and cooperation at smaller scales.

4. CONCLUSIONS

There are a number of existential risks that threaten to reduce the long-term potential of humanity. AGI and nuclear winter are two of the most important ones, but a number of different catastrophes could significantly affect agriculture (with possible long-term impact). Here the first long term future cost-effectiveness analyses is presented for AGI safety and resilient foods, a neglected, tractable intervention for agricultural catastrophes Development of the model uncovered several key areas of uncertainty.

Theoretical uncertainty represents the dominant source of uncertainty for the model. Predicting the impact of events on the long run future faces inherent epistemic uncertainty. It is acknowledged that several probability estimates contain this uncertainty, however, if one wishes to prioritize according to the long term such uncertainty must be accepted. The probability ranges of inputs concerning epistemically uncertain phenomenon are reasonable, but any readers with substantially different priors are encouraged to utilize the presented model.

Future research that would reduce model uncertainty were identified and are listed below in descending order of importance: Theoretical investigations of epistemic uncertainty relating to comparing the impact of interventions to different risks; investigating other methodologies (quantitative, qualitative) for cross risk comparisons to increase robustness to uncertainty; improved surveying and expert opinion solicitation methods e.g. Delphi method [125]; and obtaining robust cost estimates for resilient food interventions through pilot projects or planning exercises.

The epistemically reserved nature in which the presented results should be interpreted should be stressed. Model and theory uncertainty are likely large due to the inherent difficulty in modelling the complex systems discussed and the considerable epistemic challenge of assessing the long run future. There is also great parameter uncertainty in both AGI safety and resilient foods. Considering uncertainty captured within the two models presented, it can be said with 98%-99% confidence that funding resilient foods now is more cost effective than additional funding for AGI safety beyond the expected $3 billion. A sensitivity analysis was performed, and S model required three unfavorable input parameters to switch to AGI safety being more cost effective according to the mean of the distributions than resilient foods at the margin now while E model required five unfavorable input variables. Therefore, within model assumptions, the case for funding resilient foods now is robust. Based on the models, there is closer to 84%-93% confidence that spending the ~100 millionth dollar on resilient foods is more cost effective than AGI safety at the margin.

An important finding of the model is the remarkable cost effectiveness of saving lives in a 10% agricultural productivity reduction, aligning with previous cost effectiveness results of $0.20-$400 for only 10% global food production [35]. The higher likelihood of occurrence for 10% shocks between ~2x and ~ 1.3x, for S and E model respectively, and significantly lower uncertainty implies immediate preparation is justified. Such scenarios are unlikely to cause extinction but would do damage over the short term so funding from standard disaster risk reduction sources would be justified on purely short term considerations. Additionally overlap in interventions for 10% and 100% scenarios is high so preparation funding allocated for 10% would also have value for more extreme scenarios.

Resilient foods address catastrophes that have significant likelihood of occurring in the next decade, so funding is particularly urgent. Both AGI safety and resilient foods save expected lives inexpensively in the present generation, so funding should increase for both.

Disclaimer/Acknowledgements: Funding was received from the Centre for Effective Altruism. Anders Sandberg received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 669751). The Oxford Prioritisation Project developed the artificial general intelligence safety cost effectiveness submodel. Owen Cotton-Barratt, Daniel Dewey, Sindy Li, Ozzie Gooen, Kyle Alvarado, Ratheka Stormbjorne, and Finan Adamson contributed helpful discussions. Several anonymous reviewers provided useful feedback on earlier versions of the paper, and one especially rigorous reviewer whose detailed comments significantly improved the manuscript. Zvi Mowshowitz improved nuclear war probability estimates by applying Bayesian methods. This is not the official position of the Centre for Effective Altruism, the Future of Humanity Institute, nor the Alliance to Feed the Earth in Disasters (ALLFED).

APPENDIX A. Text from survey

Note: Survey refers to alternative foods. Alternative foods and resilient foods describe the same set of food production methods.

Earth-derived civilization has potential to spread to the galaxy and possibly beyond. Global catastrophes could reduce the potential of this civilization directly or with further conflict through risk of extinction, risk of losing industrial civilization and not recovering, risk of losing anthropological civilization (basically cities) and not recovering both anthropological and industrial civilization, worse human values that increase the likelihood of further global catastrophes, worse values being placed into artificial general intelligence, etc. Along with each percentage answer, please indicate your confidence level from 1 to 10. Note the numbered questions pertain to catastrophes that could collapse agriculture, while the lettered questions are on 10% global agricultural shortfalls.

1) What is your estimate of the reduction in the potential of humanity from full-scale nuclear war between the US and Russia involving thousands of nuclear weapons (if there had been no ALLFED)?

2) What is the percent mitigation of far future impact due to full-scale nuclear war due to ALLFED so far (including governments potentially finding ALLFED materials/papers on the web in a catastrophe)?

3) What is the additional percent mitigation of far future impact due to full-scale nuclear war if around $100 million were spent on planning, research, and development of alternate foods?

a) What is your estimate of the reduction in the potential of humanity from catastrophes that reduce global agricultural output by 10% suddenly (e.g. regional nuclear war like between India and Pakistan, or a volcanic eruption like the one that caused the year without a summer in 1816) (if there had been no ALLFED)?

b) What is the percent mitigation of far future impact due to 10% agricultural catastrophes due to ALLFED so far (including governments potentially finding ALLFED materials on the web in a catastrophe)?

c) What is the additional percent mitigation of far future impact due to 10% agricultural catastrophes if around $100 million were spent on planning, research, and development of alternate foods?”

APPENDIX B. Nuclear War Probabilities

S model nuclear war probability: Bayesian update of Barrett et al. 2013 probability of inadvertent nuclear war

Model S nuclear war probability is based on Barrett et al. 2013 which estimates the likelihood of inadvertent nuclear war between Russia and the United States through the development of a fault tree analysis of US nuclear systems and response procedures that have been in place since 1975 [7]:

> Systems and response procedures described here are assumed to have been used since approximately 1975, and current C3I (command, control, communication, and intelligence) systems and launch protocols have been in place for the past 37 years. There is limited publicly available data on the historical frequency of MDCs (Missile Display Conference), TACs (Threat Assessment Conference) or MACs (Missile Attack Conference) in the United States, or their equivalents in the USSR and Russia, over the same period. In the United States, during the period 1977–1983, the number of MDCs per year ranged from 43 to 255, and the number of TACs per year were either zero or two [7].

As stated above, this system's deployment since approximately 1975 means that there have been 47 years in which a failure could have occurred resulting in nuclear war. The lack of nuclear war during this period should be considered as evidence that the probability of inadvertent nuclear war is lower than estimated by the model and should be adjusted down according to Bayes theorem.

To update the Barrett et al. 2013 inadvertent nuclear war probability *P*, we assume that the non-occurrence (or occurrence) of inadvertent nuclear war is the data generating process and can represented by a Binomial distribution (1) in which each year *n* is an independent trial and inadvertent nuclear war *w* between Russia and the US is the considered event

$$P = \binom{n}{w} p^{w} p^{n-w} \qquad \text{(B1)}$$

(1) provides the likelihood function for the update. To calculate the likelihood of witnessing no instances of inadvertent nuclear war *w* = 0 in years *n = 47,* the time in years from 1975 until present (2022) a Beta prior distribution with parameters (*a* = 0.61, *β*= 30.19) is used. The Beta prior distribution parameters were determined using the *beta.select* function from the *learnbayes* Library in R [126], with the 5 and 95% confidence intervals for ‘Danger Calm’ from Barrett et al 2013 (page 120 lines 8-9) as quantiles (5% = 0.0002, 95% = 0.07). As Binomial and Beta distributions are conjugate, the posterior distribution is a Beta distribution with parameters (*a’* = 0.62, *β’* = 77.19).

Bayesian update of S and E models from 2022 to future over time horizon of effectiveness

As the model estimates the cost effectiveness of interventions over a period of time, updating the nuclear war probability for each successive year without nuclear war from present into the future is required. The Beta distribution *Beta(a’, β’ + n’)* gives the probability of inadvertent nuclear war at a n’ years into the future given no nuclear war. *n’* is a discrete integer value and is capped at the 95th percentile of the ‘time horizon of effectiveness of resilient foods’ variable for each model, i.e. 49 years for the S model and 149 years for E model after which interventions are assumed to no longer be effective due to societal and technological changes, e.g., artificial general intelligence reducing the risk or impact of nuclear war.

Updating the nuclear war probability into the future is implemented in the model by constructing an index of year intervals up to the 95% confidence interval of the time horizon of effectiveness of resilient food interventions, n_{thi} . The S Model index spans 0 - 49 years and E Model spans 0 - 149 years. The index value represents the number of years past 2022 for which no nuclear war has been observed i.e. $n' = 1,2,3,..,n_{thi}$.

Truncating probability of inadvertent nuclear war.
To further enforce sensible estimates from both models, probability of nuclear war is truncated at u = $0.95^{(1/n)}$ which is the upper limit of the 95% confidence interval for the binomial distribution containing the event probability *p* of zero successes from *n* independent trials. For the S model *n* is the number of years from 1975 (date the C3I system was implemented) until the year of the updated inadvertent nuclear war probability e.g. $n = 47 + 2 = 49$ for the inferred 2024 nuclear war probability conditional on no nuclear war between 2022 and 2024. For the E model, *n* is the number of years from 1945 until the year of the updated nuclear war probability.

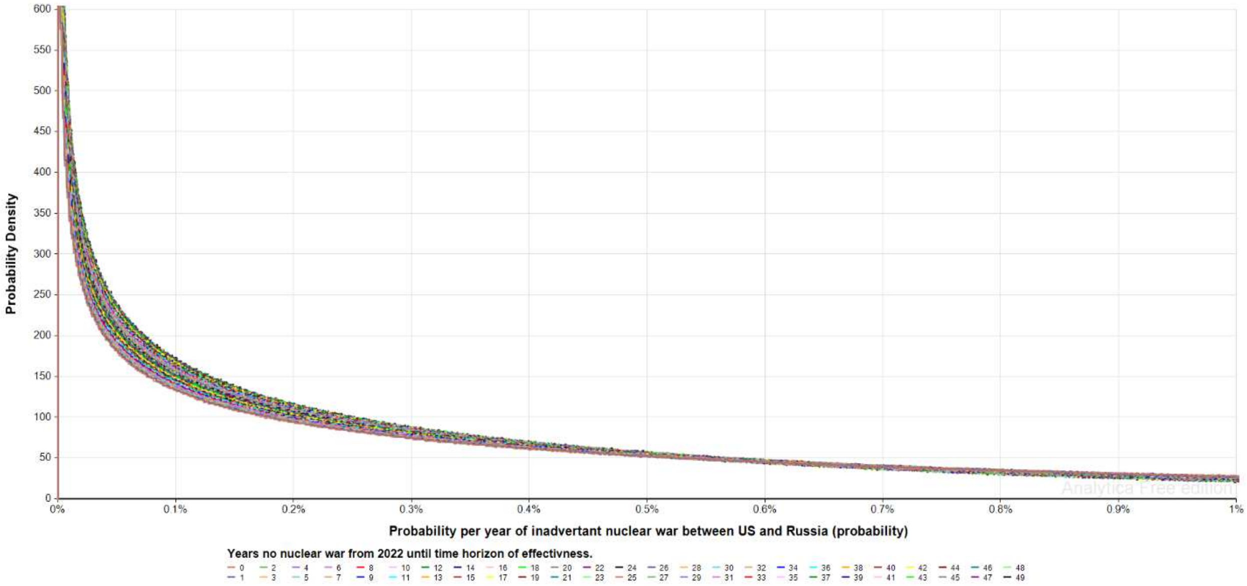


Figure B1. S model index of annual probabilities of nuclear war a given number of years from present (2022) into the future. Year zero has a fatter tail and more probability mass at higher probabilities.

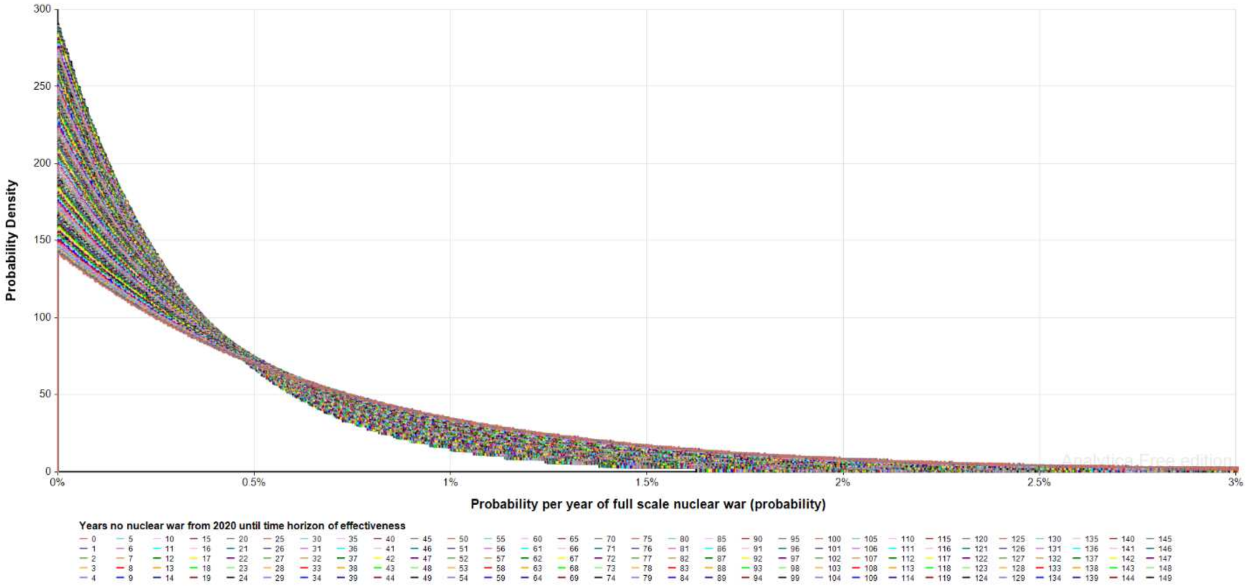


Figure B2. E model index of annual probabilities of nuclear war a given number of years from present (2022) into the future. Year zero has a fatter tail and more probability mass at higher probabilities.

Averaging values from index runs.

The model outputs an index of values corresponding to the year in the future to which nuclear war probability was updated. The values of this index are averaged in order to obtain a singular output equivalent to the average expected value.

Models are available upon request from the corresponding author.

REFERENCES

[1] J. Coupe, C.G. Bardeen, A. Robock, O.B. Toon, Nuclear Winter Responses to Nuclear War Between the United States and Russia in the Whole Atmosphere Community Climate Model Version 4 and the Goddard Institute for Space Studies ModelE, J. Geophys. Res. Atmospheres. 124 (2019) 8522–8543. https://doi.org/10.1029/2019JD030509.

[2] A. Robock, L. Oman, G.L. Stenchikov, Nuclear winter revisited with a modern climate model and current nuclear arsenals: Still catastrophic consequences: NUCLEAR WINTER REVISITED, J. Geophys. Res. Atmospheres. 112 (2007) n/a-n/a. https://doi.org/10.1029/2006JD008235.

[3] S.H. Ambrose, Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans, J. Hum. Evol. 34 (1998) 623–651. https://doi.org/10.1006/jhev.1998.0219.

[4] D.E. Fastovsky, The Extinction of the Dinosaurs in North America, GSA TODAY. (2005) 7.

[5] T. Gehrels, M.S. Matthews, A.M. Schumann, Hazards Due to Comets and Asteroids, University of Arizona Press, 1994.

[6] S. Baum, R. de Neufville, A. Barrett, A Model for the Probability of Nuclear War, SSRN Electron. J. (2018). https://doi.org/10.2139/ssrn.3137081.

[7] A.M. Barrett, S.D. Baum, K.R. Hostetler, Analyzing and reducing the risks of inadvertent nuclear war between the United States and Russia, Sci Glob. Secur. 21 (2013) 106–133.

[8] M.E. Hellman, Risk analysis of nuclear deterrence, Bent Tau Beta Pi. 99 (2008) 14.

[9] J. Reisner, G. D'Angelo, E. Koo, W. Even, M. Hecht, E. Hunke, D. Comeau, R. Bos, J. Cooley, Climate Impact of a Regional Nuclear Weapons Exchange: An Improved Assessment Based On Detailed Source Calculations, J. Geophys. Res. Atmospheres. 123 (2018) 2752–2772. https://doi.org/10.1002/2017JD027331.

[10] C.S. Lane, B.T. Chorn, T.C. Johnson, Ash from the Toba supereruption in Lake Malawi shows no volcanic winter in East Africa at 75 ka, Proc. Natl. Acad. Sci. 110 (2013) 8025–8029.

[11] A. Robock, C.M. Ammann, L. Oman, D. Shindell, S. Levis, G. Stenchikov, Did the Toba volcanic eruption of ∼74 ka B.P. produce widespread glaciation?, J. Geophys. Res. Atmospheres. 114 (2009). https://doi.org/10.1029/2008JD011652.

[12] C.L. Yost, L.J. Jackson, J.R. Stone, A.S. Cohen, Subdecadal phytolith and charcoal records from Lake Malawi, East Africa imply minimal effects on human evolution from the ∼74 ka Toba supereruption, J. Hum. Evol. 116 (2018) 75–94. https://doi.org/10.1016/j.jhevol.2017.11.005.

[13] C.P. Timmer, Reflections on food crises past, Food Policy. 35 (2010) 1–11. https://doi.org/10.1016/j.foodpol.2009.09.002.

[14] V. Smil, Energy and Civilization: A History, MIT Press, 2018.

[15] J.F. Coates, Risks and threats to civilization, humankind, and the earth, Futures. 41 (2009) 694–705. https://doi.org/10.1016/j.futures.2009.07.010.

[16] H. Greaves, W. MacAskill, The Case For Strong Longtermism, (2021). https://globalprioritiesinstitute.org/wp-content/uploads/The-Case-for-Strong-Longtermism-GPI-Working-Paper-June-2021-2-2.pdf (accessed October 23, 2021).

[17] N. Bostrom, M.M. Cirkovic, Global Catastrophic Risks, OUP Oxford, 2008.

[18] N. Bostrom, Superintelligence: paths, dangers, strategies, First edition, Oxford University Press, Oxford, 2014.

[19] P. Mangalampalli, Why can people live in Hiroshima & Nagasaki but not Chernobyl?, (2019). https://www.tutorialspoint.com/why-can-people-live-in-hiroshima-and-nagasaki-but-not-chernobyl (accessed November 22, 2019).

[20] P. Mclntyre, How you can lower the risk of a catastrophic nuclear war, 80000 Hours. (2016). https://web.archive.org/web/20190428055706/https://80000hours.org/problem-profiles/nuclear-security/ (accessed March 21, 2019).

[21] D.C. Denkenberger, J.M. Pearce, Feeding Everyone No Matter What: Managing Food Security After Global Catastrophe, Academic Press, 2014.

[22] S. Baum, D.C. Denkenberger, J.M. Pearce, Alternative foods as a solution to global food supply catastrophes, 7 (2016) 31–35. https://hal.archives-ouvertes.fr/hal-02113500.

[23] M. Abdelkhaliq, D. Denkenberger, D. Cole, M. Griswold, J. Pearce, A.R. Taylor, Non Food Needs if Industry is Disabled, in: Proc. 6th Int. Disaster Risk Conf., Davos, Switzerland, 2016.

[24] D.D. Cole, D. Denkenberger, M. Griswold, M. Abdelkhaliq, J. Pearce, Feeding Everyone if Industry is Disabled, in: IDRC DAVOS 2016 Integr. Risk Manag. - Resilient Cities, Davos, Switzerland, 2016. https://hal.archives-ouvertes.fr/hal-02113486 (accessed August 16, 2019).

[25] D.C. Denkenberger, D.D. Cole, M. Abdelkhaliq, M. Griswold, A.B. Hundley, J.M. Pearce, Feeding everyone if the sun is obscured and industry is disabled, Int. J. Disaster Risk Reduct. 21 (2017) 284–290. https://doi.org/10.1016/j.ijdrr.2016.12.018.

[26] J. Throup, B. Bals, J. Cates, J.B. García Martínez, J.M. Pearce, D.C. Denkenberger, Rapid Repurposing of Biorefinery, Pulp & Paper and Breweries for Lignocellulosic Sugar Production in Global Food Shortages, (2020). https://doi.org/10.31219/osf.io/jns2e.

[27] J.B. García Martínez, J. Egbejimba, J. Throup, S. Matassa, J.M. Pearce, D.C. Denkenberger, Potential of microbial protein from hydrogen for preventing mass starvation in catastrophic scenarios, Sustain. Prod. Consum. 25 (2021) 234–247. https://doi.org/10.1016/j.spc.2020.08.011.

[28] K.A. Alvarado, J.B. García Martínez, S. Matassa, J. Egbejimba, D. Denkenberger, Food in space from hydrogen-oxidizing bacteria, Acta Astronaut. 180 (2021) 260–265. https://doi.org/10.1016/j.actaastro.2020.12.009.

[29] J.B. García Martínez, M.M. Brown, X. Christodoulou, K.A. Alvarado, D.C. Denkenberger, Potential of microbial electrosynthesis for contributing to food production using CO2 during global agriculture-inhibiting disasters, Clean. Eng. Technol. 4 (2021). https://doi.org/10.1016/j.clet.2021.100139.

[30] J.B. García Martínez, K.A. Alvarado, X. Christodoulou, D.C. Denkenberger, Chemical synthesis of food from CO2 for space missions and food resilience, J. CO2 Util. 53 (2021). https://doi.org/10.1016/j.jcou.2021.101726.

[31] A. Mill, C. Harrison, S. James, S. Shah, T. Fist, K. Alvarado, A. Taylor, D. Denkenberger, Preventing global famine in case of sun-blocking scenarios: Seaweed as an alternative food source, in: 2019. https://www.researchgate.net/publication/337199859_Preventing_global_famine_in_case_of_sun-blocking_scenarios_Seaweed_as_an_alternative_food_source_Key_findings.

[32] K.A. Alvarado, A. Mill, J.M. Pearce, A. Vocaet, D. Denkenberger, Scaling of greenhouse crop production in low sunlight scenarios, Sci. Total Environ. 707 (2020) 136012. https://doi.org/10.1016/j.scitotenv.2019.136012.

[33] D.C. Denkenberger, J.M. Pearce, Micronutrient Availability in Alternative Foods During Agricultural Catastrophes, Agriculture. 8 (2018) 169. https://doi.org/10.3390/agriculture8110169.

[34] M. Griswold, D. Denkenberger, M. Abdelkhaliq, D. Cole, J. Pearce, A.R. Taylor, Vitamins in Agricultural Catastrophes, in: Proc. 6th Int. Disaster Risk Conf., Davos, Switzerland, 2016.

[35] D.C. Denkenberger, J.M. Pearce, Cost-Effectiveness of Interventions for Alternate Food to Address Agricultural Catastrophes Globally, Int. J. Disaster Risk Sci. 7 (2016) 205–215. https://doi.org/10.1007/s13753-016-0097-2.

[36] D. Denkenberger, J. Pearce, A.R. Taylor, R. Black, Food without sun: price and life-saving potential, Foresight. 21 (2019) 118–129. https://doi.org/10.1108/FS-04-2018-0041.

[37] P. Valdes, Built for stability, Nat Geosci. 4 (2011) 414–416. https://doi.org/10.1038/ngeo1200.

[38] J.P. Dudley, M.H. Woodford, Bioweapons, Biodiversity, and Ecocide: Potential Effects of Biological Weapons on Biological Diversity, BioScience. 52 (2002) 583. https://doi.org/10.1641/0006-3568(2002)052[0583:BBAEPE]2.0.CO;2.

[39] C.C. Mann, Genetic engineers aim to soup up crop photosynthesis, Sci. 283 (1999) 314–316.

[40] H. Saigo, Agricultural Biotechnology and the Negotiation of the Biosafety Protocol, Georget. Int. Environ. Law Rev. 12 (1999) 779.

[41] S. Dietz, High impact, low probability? An empirical analysis of risk in the economics of climate change, Clim. Change. 108 (2011) 519–541. https://doi.org/10.1007/s10584-010-9993-4.

[42] D.C. Denkenberger, J.M. Pearce, A National Pragmatic Safety Limit for Nuclear Weapon Quantities, Safety. 4 (2018) 25. https://doi.org/10.3390/safety4020025.

[43] M.J. Cohen, M. Smale, Global food-price shocks and poor people - an overview, Dev. Pract. 21 (2011) 460–471.

[44] T. Ord, The Precipice: Existential Risk and the Future of Humanity, Hachette Books, 2020.

[45] T. Schaul, J. Togelius, J. Schmidhuber, Measuring intelligence through games, ArXiv Prepr. ArXiv11091314. (2011).

[46] R. Dale, GPT-3: What's it good for?, Nat. Lang. Eng. 27 (2021) 113–118. https://doi.org/10.1017/S1351324920000601.

[47] I.J. Good, Speculations concerning the first ultraintelligent machine, in: Adv. Comput., Elsevier, 1966: pp. 31–88.

[48] N. Bostrom, The Superintelligent Will: Motivation and Instrumental Rationality in Advanced Artificial Agents, Minds Mach. 22 (2012) 71–85. https://doi.org/10.1007/s11023-012-9281-3.

[49] B. Todd, Why despite global progress, humanity is probably facing its most dangerous time ever, 80000 Hours. (2017). https://80000hours.org/articles/extinction-risk/ (accessed August 15, 2020).

[50] LessWrong, Existential risk, (2020). https://wiki.lesswrong.com/wiki/Existential_risk (accessed August 15, 2020).

[51] A. Turchin, D. Denkenberger, Classification of global catastrophic risks connected with artificial intelligence, AI Soc. (2018). https://doi.org/10.1007/s00146-018-0845-5.

[52] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, D. Mané, Concrete Problems in AI Safety, ArXiv160606565 Cs. (2016). http://arxiv.org/abs/1606.06565 (accessed April 11, 2019).

[53] J.G. Matheny, Reducing the Risk of Human Extinction, Risk Anal. 27 (2007) 1335–1344. https://doi.org/10.1111/j.1539-6924.2007.00960.x.

[54] J. Halstead, Climate Change Cause Area Report, (2018).

[55] P. Millett, A. Snyder-Beattie, Existential Risk and Cost-Effective Biosecurity, Health Secur. 15 (2017) 373–383. https://doi.org/10.1089/hs.2017.0028.

[56] B.J. Garrick, Quantifying and controlling catastrophic risks, Academic Press, 2008.

[57] S. Li, A model of the Machine Intelligence Research Institute, Oxf. Prioritisation Proj. (2017). https://oxpr.io/blog/2017/5/20/a-model-of-the-machine-intelligence-research-institute (accessed August 13, 2020).

[58] Oxford Prioritisation Project, Machine Intelligence Research Institute - Oxford Prioritisation Project, Guesstimate. (2017). https://www.getguesstimate.com/models/8789 (accessed June 18, 2021).

[59] A. Sandberg, D. Manheim, What is the Upper Limit of Value?, 2021.

[60] M. Keramat, R. Kielbasa, Latin hypercube sampling Monte Carlo estimation of average quality index for integrated circuits, in: Analog Des. Issues Digit. VLSI Circuits Syst., Springer, 1997: pp. 131–142.

[61] D. Denkenberger, O. Cotton-Barrat, D. Dewey, S. Li, Food without the sun and AI X risk cost effectiveness general far future impact publication, Guesstimate. (2019). https://www.getguesstimate.com/models/13082 (accessed April 11, 2019).

[62] D.A. Raitzer, T.G. Kelley, Benefit-cost meta-analysis of investment in the International Agricultural Research Centers of the CGIAR, Agric. Syst. 96 (2008) 108–123. https://doi.org/10.1016/j.agsy.2007.06.004.

[63] CGIAR, CGIAR - Financial Report 2009, (2009). https://web.archive.org/web/20111103163833/http://www.cgiar.org/publications/annual/cgiar-annualreport2010/cgiar-annualreport2010/cgiar_annual_report/2009_CGIAR_Full_Financial_Report_FINAL.pdf.pdf (accessed October 21, 2021).

[64] A. Aierzhati, J. Watson, B. Si, M. Stablein, T. Wang, Y. Zhang, Development of a mobile, pilot scale hydrothermal liquefaction reactor: Food waste conversion product analysis and techno-economic assessment, Energy Convers. Manag. X. 10 (2021) 100076. https://doi.org/10.1016/j.ecmx.2021.100076.

[65] M. Villain-Gambier, M. Courbalay, A. Klem, S. Dumarcay, D. Trebouet, Recovery of lignin and lignans enriched fractions from thermomechanical pulp mill process water through membrane separation technology: Pilot-plant study and techno-economic assessment, J. Clean. Prod. 249 (2020) 119345. https://doi.org/10.1016/j.jclepro.2019.119345.

[66] K.F. Lam, C.C.J. Leung, H.M. Lei, C.S.K. Lin, Economic feasibility of a pilot-scale fermentative succinic acid production from bakery wastes, Food Bioprod. Process. 92 (2014) 282–290. https://doi.org/10.1016/j.fbp.2013.09.001.

[67] N. Beintema, A.N. Pratt, G.-J. Stads, KEY TRENDS IN GLOBAL AGRICULTURAL RESEARCH INVESTMENT, (n.d.) 8.

[68] A. Grubler, The Rise and Fall of Infrastructures: Dynamics of Evolution and Technological Change in Transport, Physica-Verlag, Heidelberg, 1990. http://pure.iiasa.ac.at/id/eprint/3351/ (accessed August 15, 2020).

[69] A. Grübler, Time for a Change: On the Patterns of Diffusion of Innovation, Daedalus. 125 (1996) 19–42.

[70] E.W. Montroll, Social dynamics and the quantifying of social forces, Proc. Natl. Acad. Sci. U. S. A. 75 (1978) 4633–4637.

[71] G. Allison, Destined for War: Can America and China Escape Thucydides's Trap?, Houghton Mifflin Harcourt, 2017.

[72] R. Bailey, T.G. Benton, A. Challinor, J. Elliott, D. Gustafson, B. Hiller, A. Jones, C. Kent, K. Lewis, T. Meacham, M. Rivington, R. Tiffin, D.J. Wuebbles, Extreme weather and resilience of the global food system: Final Project Report from the UK-US Taskforce on Extreme Weather and Global Food System Resilience, UK Glob. Food Secur. Programme. (2015). https://www.stat.berkeley.edu/~aldous/157/Papers/extreme_weather_resilience.pdf.

[73] R. Duda, Report: Is climate change the biggest problem in the world?, 80000 Hours. (2016). https://web.archive.org/web/20190203015300/https://80000hours.org/problem-profiles/climate-change/#fn-ref-1 (accessed April 11, 2019).

[74] E. Rindzevičiūtė, FROM NUCLEAR WINTER TO THE ANTHROPOCENE, in: Power Syst., Cornell University Press, 2016: pp. 150–180. https://www.jstor.org/stable/10.7591/j.ctt1d2dmw5.11 (accessed April 11, 2019).

[75] O.B. Toon, A. Robock, R.P. Turco, Environmental consequences of nuclear war, Phys. Today. 61 (2008) 37–42. https://doi.org/10.1063/1.3047679.

[76] H.M. Kristensen, R.S. Norris, Chinese nuclear forces, 2018, Bull. At. Sci. 74 (2018) 289–295. https://doi.org/10.1080/00963402.2018.1486620.

[77] N. Bostrom, Ethical Issues In Advanced Artificial Intelligence, (2003). https://nickbostrom.com/ethics/ai.html (accessed August 15, 2020).

[78] E. Yudkowsky, Artificial Intelligence as a Positive and Negative Factor in Global Risk, (2008) 46.

[79] P. Christiano, What failure looks like - LessWrong 2.0, (2019). https://www.lesswrong.com/posts/HBxe6wdjxK239zajf/what-failure-looks-like (accessed August 13, 2020).

[80] K. Sotala, Disjunctive Scenarios of Catastrophic AI Risk, Artif. Intell. Saf. Secur. (2018). https://doi.org/10.1201/9781351251389-22.

[81] N. Bostrom, The Vulnerable World Hypothesis, Glob. Policy. 10 (2019) 455–476. https://doi.org/10.1111/1758-5899.12718.

[82] M.J. Rees, Our Final Hour: A Scientist's Warning : how Terror, Error, and Environmental Disaster Threaten Humankind's Future in this Century--on Earth and Beyond, Basic Books, 2003.

[83] F. Simpson, Apocalypse Now? Reviving the Doomsday Argument, ArXiv161103072 Phys. Stat. (2016). http://arxiv.org/abs/1611.03072 (accessed July 8, 2021).

[84] Will humans go extinct by 2100?, (2017). https://www.metaculus.com/questions/578/human-extinction-by-2100/ (accessed July 8, 2021).

[85] O. Cotton-Barratt, The law of logarithmic returns, Future Humanity Inst. (2014). http://www.fhi.ox.ac.uk/law-of-logarithmic-returns/ (accessed April 11, 2019).

[86] ALLFED, Home, ALLFED. (2020). http://allfed.info/ (accessed April 11, 2019).

[87] S. Nellis, Microsoft to invest $1 billion in OpenAI, Reuters. (2019). https://www.reuters.com/article/us-microsoft-openai-idUSKCN1UH1H9 (accessed June 11, 2021).

[88] Open Philanthropy, Grants Database, Open Philanthr. (2021). https://www.openphilanthropy.org/giving/grants (accessed June 9, 2021).

[89] 2019 AI Alignment Literature Review and Charity Comparison - LessWrong, (n.d.). https://www.lesswrong.com/posts/SmDziGM9hBjW9DKmf/2019-ai-alignment-literature-review-and-charity-comparison (accessed June 11, 2021).

[90] A.M. Barrett, Value of GCR Information: Cost Effectiveness-Based Approach for Global Catastrophic Risk (GCR) Reduction, Forthcom. Decis. Anal. (2017).

[91] O. Cotton-Barratt, M. Daniel, A. Sandberg, Defence in Depth Against Human Extinction: Prevention, Response, Resilience, and Why They All Matter, Glob. Policy. 11 (2020) 271–282. https://doi.org/10.1111/1758-5899.12786.

[92] National Nuclear Security Administration, FY 2021 Presidential Budget for NNSA Released, Energy.Gov. (2020). https://www.energy.gov/nnsa/budget (accessed August 13, 2020).

[93] Nuclear Weapons Policy, Nucl. Weapons Policy. (2019). 9:57 pm 8/12/2020.

[94] Global Catastrophic Risk Institute, Nuclear War, Glob. Catastrohpic Risk Inst. (n.d.). http://gcrinstitute.org/nuclear/.

[95] Effective Altruism Concepts, Importance, tractability, neglectedness framework, Eff. Altruism Concepts. (2019). https://concepts.effectivealtruism.com/concepts/importance-neglectedness-tractability/ (accessed April 11, 2019).

[96] Calysta Inc., Adisseo and Calysta establish a Joint-Venture to commercialize FeedKind®, (2020). http://www.feedkind.com/adisseo-calysta-establish-joint-venture-commercialize-feedkind/ (accessed May 24, 2020).

[97] S.W. Jones, A. Karpol, S. Friedman, B.T. Maru, B.P. Tracy, Recent advances in single cell protein use as a feed ingredient in aquaculture, Curr. Opin. Biotechnol. 61 (2020) 189–197. https://doi.org/10.1016/j.copbio.2019.12.026.

[98] E. Penrod, Companies plan launch, expansion of single-cell proteins, Feed Strategy. (2021). https://www.feedstrategy.com/animal-feed-additives/companies-plan-launch-expansion-of-single-cell-proteins/ (accessed July 10, 2021).

[99] GiveWell, Cost-Effectiveness, GiveWell. (2017). https://www.givewell.org/how-we-work/our-criteria/cost-effectiveness (accessed April 11, 2019).

[100] T. Ord, The timing of labour aimed at reducing existential risk, Future Humanity Inst. (2014). https://www.fhi.ox.ac.uk/the-timing-of-labour-aimed-at-reducing-existential-risk/ (accessed April 11, 2019).

[101] S. Farquhar, Changes in funding in the AI safety field, Eff. Altruism. (2017). https://www.effectivealtruism.org/articles/changes-in-funding-in-the-ai-safety-field/ (accessed April 11, 2019).

[102] S. Islam, K.M. Zobair, C. Chu, J.C.R. Smart, M.S. Alam, Do Political Economy Factors Influence Funding Allocations for Disaster Risk Reduction?, J. Risk Financ. Manag. 14 (2021) 85. https://doi.org/10.3390/jrfm14020085.

[103] C. Mom, U. Sandström, P. van den Besselaar, Does cronyism affect grant application success? The role of organizational proximity, STI 2018 Conf. Proc. (2018) 1579–1585.

[104] A. Ebadi, A. Schiffauerova, How to Receive More Funding for Your Research? Get Connected to the Right People!, PLOS ONE. 10 (2015) e0133061. https://doi.org/10.1371/journal.pone.0133061.

[105] W.P. Wahls, High cost of bias: Diminishing marginal returns on NIH grant funding to institutions, 2018. https://doi.org/10.1101/367847.

[106] S. Dickert, D. Västfjäll, J. Kleber, P. Slovic, Scope insensitivity: The limits of intuitive valuation of human lives in public policy, J. Appl. Res. Mem. Cogn. 4 (2015) 248–255. https://doi.org/10.1016/j.jarmac.2014.09.002.

[107] M. Cassidy, L. Mani, On the assessment of volcanic eruptions as global catastrophic or existential risks, (n.d.). https://forum.effectivealtruism.org/posts/jJDuEhLpF7tEThAHy/on-the-assessment-of-volcanic-eruptions-as-global (accessed October 19, 2021).

[108] M. Henchion, M. Hayes, A.M. Mullen, M. Fenelon, B. Tiwari, Future protein supply and demand: strategies and factors influencing a sustainable equilibrium, Foods. 6 (2017) 53. https://doi.org/10.3390/foods6070053.

[109] E. Mahembe, N.M. Odhiambo, Foreign aid and poverty reduction: A review of international literature, Cogent Soc. Sci. 5 (2019) 1625741. https://doi.org/10.1080/23311886.2019.1625741.

[110] Development Initiatives, Global Humanitarian Assistance Report 2021, Development Initiatives, 2021. https://devinit.org/resources/global-humanitarian-assistance-report-2021/ (accessed October 19, 2021).

[111] T. Ord, R. Hillerbrand, A. Sandberg, Probing the improbable: methodological challenges for risks with low probabilities and high stakes, J. Risk Res. 13 (2010) 191–205. https://doi.org/10.1080/13669870903126267.

[112] H. Greaves, Cluelessness, Proc. Aristot. Soc. 116 (2016) 311–339. https://doi.org/10.1093/arisoc/aow018.

[113] N. Bostrom, Existential Risk Prevention as Global Priority, Glob. Policy. 4 (2013) 15–31. https://doi.org/10.1111/1758-5899.12002.

[114] N. Bostrom, Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards, J. Evol. Technol. 9 (2002). https://www.nickbostrom.com/existential/risks.pdf (accessed August 22, 2020).

[115] S.D. Baum, The far future argument for confronting catastrophic threats to humanity: Practical significance and alternatives, Futures. 72 (2015) 86–96. https://doi.org/10.1016/j.futures.2015.03.001.

[116] C. Tarsney, The Epistemic Challenge to Longtermism, (2020).

[117] M.T. Orne, On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications., in: 1962. https://doi.org/10.1037/h0043424.

[118] D. Kahneman, Thinking Fast and Slow, New York :Farrar, Straus and Giroux, 2011.

[119] J. Shanteau, D.J. Weiss, R.P. Thomas, J.C. Pounds, Performance-based assessment of expertise: How to decide if someone is an expert or not, Eur. J. Oper. Res. 136 (2002) 253–263. https://doi.org/10.1016/S0377-2217(01)00113-8.

[120] A.L. Koch, The logarithm in biology 1. Mechanisms generating the log-normal distribution exactly, J. Theor. Biol. 12 (1966) 276–290. https://doi.org/10.1016/0022-5193(66)90119-6.

[121] E. Limpert, W.A. Stahel, M. Abbt, Log-normal Distributions across the Sciences: Keys and CluesOn the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can

provide deeper insight into variability and probability—normal or log-normal: That is the question, BioScience. 51 (2001) 341–352. https://doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2.

[122] M. Mitzenmacher, A Brief History of Generative Models for Power Law and Lognormal Distributions, Internet Math. 1 (2003) 226–251.

[123] L. Chrisman, M. Henrion, R. Morgan, B. Arnold, F. Brunton, A. Esztergar, J. Harlan, Analytica user guide, Lumina Decision Systems, Los Gatos, CA, 2007.

[124] M.G. Morgan, M. Henrion, Uncertainty: a Guide to dealing with uncertainty in quantitative risk and policy analysis Cambridge University Press, N. Y. N. Y. USA. (1990).

[125] S. Beard, T. Rowe, J. Fox, An analysis and evaluation of methods currently used to quantify the likelihood of existential hazards, Futures. 115 (2020) 102469. https://doi.org/10.1016/j.futures.2019.102469.

[126] J. Albert, Functions for Learning Bayesian Inference, (2018). https://cran.r-project.org/web/packages/LearnBayes/LearnBayes.pdf (accessed February 4, 2022).