Nuclear war tail risk has been exaggerated?

The views expressed here are my own, not those of Alliance to Feed the Earth in Disasters (ALLFED), for which I work as a contractor.

Summary

- I calculated a nearterm annual risk of human extinction from nuclear war of 5.93*10^-12 (more).
- I consider grantmakers and donors interested in decreasing extinction risk had better focus on artificial intelligence (AI) instead of nuclear war (more).
- I would say the case for sometimes prioritising nuclear extinction risk over AI extinction risk is much weaker than the case for sometimes prioritising natural extinction risk over nuclear extinction risk (more).
- I get a sense the extinction risk from nuclear war was massively overestimated in The Existential Risk Persuasion Tournament (XPT) (more).
- I have the impression Toby Ord greatly overestimated tail risk in The Precipice (more).
- I believe interventions to decrease deaths from nuclear war should be assessed based on standard cost-benefit analysis (more).
- I think increasing calorie production via new food sectors is less cost-effective to save lives than measures targeting distribution (more).

Extinction risk from nuclear war

I calculated a nearterm annual risk of human extinction from nuclear war of 5.93*10^-12 (= (6.36*10^-14*5.53*10^-10)^0.5) from the geometric mean between[1]:

- My prior of 6.36*10^-14 for the annual probability of a war causing human extinction.
- My inside view estimate of 5.53*10^-10 for the nearterm annual probability of human extinction from nuclear war.

By nearterm annual risk, I mean that in a randomly selected year from 2025 to 2050. I computed my inside view estimate of 5.53*10^-10 (= 0.0131*0.0422*10^-6) multiplying:

- 1.31 % annual probability of a nuclear weapon being detonated as an act of war.
- 4.22 % probability of insufficient calorie production given at least one nuclear detonation.
- 10^-6 probability of human extinction given insufficient calorie production.

I explain the rationale for the above estimates in the next sections. Note nuclear war might have cascade effects which lead to civilisational collapse[2], which could increase longterm

[1] The geometric mean between 2 small probabilities is similar to the probability linked to the geometric mean of the odds of the 2 probabilities.

[2] For instance, Bailey 2017 analyses the effects of interruptions at chokepoints in global food trade. "Critical junctures on transport routes through which exceptional volumes of trade pass". A reviewer

extinction risk while simultaneously having a negligible impact on the nearterm one I estimated. I do not explicitly assess this in the post, but I guess the nearterm annual risk of human extinction from nuclear war is a good proxy for the importance of decreasing nuclear risk from a longtermist perspective:

- My prior implicitly accounts for the cascade effects of wars. I derived it from historical data on the deaths of combatants due to not only fighting, but also disease and starvation, which are ever-present indirect effects of war.
- Nuclear war might have cascade effects, but so do other catastrophes.
- Global civilisational collapse due to nuclear war seems very unlikely to me. For instance, the maximum destroyable area by any country in a nuclear 1st strike was estimated to be 65.3 k km^2 in Suh 2023 (for a strike by Russia), which is just 70.8 % (= 65.3*10^3/(92.2*10^3)) of the area of Portugal, or 3.42 % (= 65.3*10^3/(1.91*10^6)) of the global urban area.
- Even if nuclear war causes a global civilisational collapse which eventually leads to extinction, I guess full recovery would be extremely likely. In contrast, an extinction caused by advanced AI would arguably not allow for a full recovery.
- I am open to the idea that nuclear war can have longterm implications even in the case of full recovery, but considerations along these lines would arguably be more pressing in the context of AI risk.
    - For context, William MacAskill said the following on The 80,000 Hours Podcast. “It’s quite plausible, actually, when we look to the very long-term future, that that’s [whether artificial general intelligence is developed in “liberal democracies” or “in some dictatorship or authoritarian state”] the biggest deal when it comes to a nuclear war: the impact of nuclear war and the distribution of values for the civilisation that returns from that, rather than on the chance of extinction”.
    - Nevertheless, value lock-in (for better or worse) is arguably more cost-effectively ensured via influencing the development of AI.
- Appealing to cascade effects or other known unknowns feels a little like a regression to the inscrutable, which is characterised by the following pattern:
    - Arguments for high existential risk initially focus on aspects of the risk which are relatively better understood (e.g. famine deaths due to the climatic effects of nuclear war).
    - Further analysis frequently shows the risk from such aspects has been overestimated, and is in fact quite low (e.g. nearterm risk of human extinction from nuclear war).
    - Then discussions move to more poorly understood aspects of the risk (e.g. how the distribution of values after a nuclear war affects the longterm values of transformative AI).

In any case, I recognise it is a crucial consideration whether nearterm annual risk of human extinction from nuclear war is a good proxy for the importance of decreasing nuclear risk from a longtermist perspective. I would agree further research on this is really valuable.

---

highlighted other cascade effects which might lead to civilisational collapse: loss of major world governments, major changes in the distribution military of power; loss of power grids, fuel supply chains, and many machines and devices through direct destruction and nuclear electromagnetic pulses (nuclear EMPs); loss of major nodes in the financial and transportation system; uncontrolled wildfires; and further crop and animal losses from radiation.

Additionally, I appreciate one should be sceptical whenever a model outputs a risk as low as the ones I mentioned at the start of this section. For example, a model predicting a 1 in a trillion chance of the global real gross domestic product (real GDP) decreasing from 2008 to 2009 would certainly not be capturing most of the actual risk of recession then, which would come from that model being (massively) wrong. On the other hand, one should be careful not to overgeneralise this type of reasoning, and conclude that any model outputting a small probability must be wrong by many orders of magnitude (OOMs). The global real GDP decreased 0.743 % (= 1 - 92.21/92.9) from 2008 to 2009, largely owing to the 2007–2008 financial crisis, but such a tiny drop is a much less extreme event than human extinction. Basic analysis of past economic trends would have revealed global recessions are unlikely, but perfectly plausible. In contrast, I see historical data suggesting a war causing human extinction is astronomically unlikely.

Finally, one could claim I am underestimating the risk due to not adequately accounting for unknown unknowns. I agree, but:

- I might as well be overestimating it for the same reasons. To illustrate, one knows nothing about absolutely unknown unknowns, and therefore should not expect them to move the best guess for the risk up or down[3].
- In the real world of probabilities, if not in that of logic, absence of evidence is evidence of absence.
- I have the impression best guesses for tail risk and cost-effectiveness usually go down[4].
- It is harder to decrease the risks from unknown unknowns because there is less information about them.
- Unknown unknowns also affect other risks, and it is unclear whether the unknown unknowns surrounding nuclear and AI risk are such that I am underestimating the importance of the former relative to the latter.

Annual probability of a nuclear weapon being detonated as an act of war

I estimated an annual probability of a nuclear weapon being detonated as an act of war of 1.31 % (= 1 - (1 - 0.29)^(1/(2050 - 2024))), which I got from Metaculus' community prediction on 23 January 2024 of 29 % before 2050. My annual probability is 1.03 (= 0.0131/0.0127) times the base rate of 1.27 % (= 1/79), respecting nuclear detonations in one year over the last 79 (= 2023 - 1945 + 1), which seems reasonable.

[3] In reality, people use the term unknown unknowns to refer to considerations about which we have some understanding.

[4] Note that best guesses going down is often weak evidence that they were overestimates. A best guess should in expectation stay the same, but this is compatible with it being more likely to go down than up. The expected value of a heavy-tailed distribution can be much larger than its median, so it can be quite likely that one's best, respecting the expected value, goes down as one updates towards a distribution with less uncertainty.

Probability of insufficient calorie production given at least one nuclear detonation

I determined a probability of (globally) insufficient calorie production given at least one nuclear detonation of 4.22 %. I computed this running a Monte Carlo simulation with 1 M samples and independent distributions[5], and supposing:

- The number of nuclear detonations given at least one being detonated as an act of war, as a fraction of the total of 12.5 k in 2023, is described by a beta distribution with 61st percentile (= 1 - 0.39) of 0.800 % (= 100/(12.5*10^3)), and 89th percentile (= 1 - 0.11) of 8.00 % (= 1*10^3/(12.5*10^3)), which has alpha and beta parameters of 0.190 and 6.68, and mean of 2.77 %. The 61st and 89th percentiles correspond to Metaculus' community predictions on 2 February 2024 of 39 % and 11 % probability of over 100 and 1 k offensive nuclear detonations before 2050 given at least one nuclear detonation causing a fatality before 2050.
- The fraction of nuclear detonations which are countervalue[6] is represented by a beta distribution with 25th and 75th percentiles equal to 3.7 % and 63.0 %, in agreement with Metaculus' community predictions on 2 February 2024. This beta distribution has alpha and beta parameters of 0.364 and 0.682, and mean of 34.8 %.
- The mean equivalent yield of the countervalue nuclear detonations is 121 kt[7] (= 2,559*10^6/21,234), which I got from the ratio between:
    - 2,559 Mt (= 1,261 + 1,006 + 167 + 74 + 31 + 14 + 6) equivalent yield deliverable in a nuclear 1st strike in 2010[8], summed across countries.

---

[5] The running time is 0.5 s.

[6] Jeffrey Lewis clarified on The 80,000 Hours Podcast there is not a sharp distinction between counterforce and countervalue:
"And so just to explain that a little bit, or unpack that: if you look at what the United States says about its nuclear weapons today, we are explicit that we target things that the enemy values, and we are also explicit that we follow certain interpretations of the law of armed conflict. And it is absolutely clear in those legal writings that the United States does not target civilians intentionally, but that in conducting what you might call "counterforce," there is a list of permissible targets. And they include not just nuclear forces. I think often in the EA community, people assume counterforce means nuclear forces, because it's got the word "force," right? But it's not true. So traditionally, the US targets nuclear forces and all of the supporting infrastructure — including command and control, it targets leadership, it targets other military forces, and it targets what used to be called "war-supporting industries," but now are called "war-sustaining industries.""

In the context of the Metaculus' prediction:
"A strike is considered countervalue for these purposes if credible media reporting does not widely consider a military or industrial target as the primary target of the attack (except in the case of strikes on capital cities, which will automatically be considered countervalue for this question even if credible media report that the rationale for the strike was disabling command and control structures)."

[7] I did not model this as a distribution because its uncertainty is much smaller than that in other factors for the cases I am interested in (relatedly). I am analysing extinction risk, so I want the distribution to be accurate for cases with many detonations. Since the mean equivalent yield tends to a constant as the detonations tend to the available nuclear warheads, I think using a constant is appropriate. In addition, the importance of modelling a factor in a product as a distribution decreases with the number of factors which are already being modelled as a distribution. If N factors follow independent lognormal distributions whose ratio between the 95th and 5th percentile is r, the ratio between the 95th and 5th percentile of the distribution of the product is r^(N^0.5). The exponent grows sublinearly with the number of factors, so the relative increase in the uncertainty of the product is smaller if one is already modelling many of its factors as distributions.

[8] The equivalent yield is defined in Suh 2023 such that it is proportional to the destroyable area. From equations 1 and 2, the equivalent yield is proportional to the yield to the power of 2/3 if the yield is

    - 21,234 nuclear warheads in 2010.
- The soot injected into the stratosphere per equivalent yield is the maximum likelihood lognormal distribution given 2 independent estimates of 3.15*10^-5 and 0.00215 Tg/kt.
    - I arrived at these by adjusting results from Reisner 2018 and Reisner 2019, and Toon 2008 and Toon 2019.
    - The mean and standard deviation of the logarithm of the distribution I just mentioned are equal to the mean and unadjusted standard deviation of the logarithms of the 2 estimates, which are -8.25 and 2.11[9].
    - For references, my mean soot injected into the stratosphere per equivalent yield is 0.00242 Tg/kt, which is 1.13 (= 0.00242/0.00215) times my higher estimate. The reasons for this are the distribution having to be quite wide for one to be maximally likely to observe 2 very different estimates, and the mean of a lognormal distribution increasing with its uncertainty[10].
- Minimum soot injected into the stratosphere for insufficient calorie production of 84.2 Tg (= 47 + (150 - 47)/(2.38 - 1.08)*(2.38 - 1.91)). This is the minimum for insufficient calorie consumption in year 2[11], less than 1.91 k kcal/person/d, given equitable food distribution, consumption of all edible livestock feed, and no household food waste, linearly interpolating the data of Fig. 5a of Xia 2022:
    - 47 Tg for 2.38 k kcal/person/d[12].
    - 150 Tg for 1.08 k kcal/person/d[13].
- The net effect on calorie production of all the adaptation measures is similar to assuming equitable food distribution, consumption of all edible livestock feed, and no household food waste. To the extent these 3 are needed to mitigate famine nationally, I guess they would be roughly fully implemented nationally, but not globally. Nevertheless, there are other factors contributing towards Xia 2022 overestimating famine (relatedly, see resilient food solutions):
    - The baseline conditions in Xia 2022 refer to 2010, but the world is becoming increasingly more resilient against starvation. The death rate from protein-energy malnutrition decreased 77.7 % (= 1 - (0.00274 %)/(0.0123 %)) from 1990 to 2019[14].

---

smaller than 1 Mt, and to the yield to the power of 1/2 if the yield is larger than 1 Mt. I actually think the (maximum) burnable area is proportional to the yield, thus being larger than the destroyable area estimated in Suh 2023. On the other hand, the actual burned area will be smaller than the burnable area, which counteracts the effect of using a higher exponent of 1. In any case, using an exponent of 1 instead of 2/3 to estimate the equivalent yield only makes the burnable area 1.14 times as large for the nuclear arsenal of the United States in 2023. So I guess the question of which exponent to use is not that important, especially in the context of estimating extinction risk.

[9] By unadjusted standard deviation, I mean the square root of the unadjusted variance.

[10] The mean of a lognormal distribution can be expressed as m*e^(sigma^2/2), where m is the median of the lognormal, and sigma is the standard deviation of the logarithm of the lognormal.

[11] In Xia 2022, “the soot is arbitrarily injected during the week starting on May 15 of Year 1”.

[12] I obtained high precision based on the pixel coordinates of the relevant points, which I retrieved with Paint.

[13] I obtained high precision based on the pixel coordinates of the relevant points, which I retrieved with Paint.

[14] Interestingly, the annual FAO Food Price Index (FFPI), which “is a measure of the monthly [and annual] change in international prices of a basket of food commodities”, increased 51.0 % (= 95.1/63.0 - 1) during the same period (calculated based on values in column B of tab “Annual” of the excel file “Excel: Nominal and real indices from 1990 onwards (monthly and annual)”). So the FFPI is not a good proxy for the death rate from protein-energy malnutrition. I believe this is explained by

- Foreign aid to the more affected countries, including international food assistance.
- Increase in meat production per capita from 2010, which is the reference year in Xia 2022.
- Increase in real GDP per capita from 2010, which is relevant because poverty is a major risk factor for famines.
- Replacing forest and grazing land by cropland:
  - In 2016, grazing land was 2.06 (= 3.28/1.59) times as large as cropland, so this would become 3.06 (= 1 + 2.06) times as large given full replacement.
  - In 2019, forest land was 85.5 % (= 0.3758/0.4394) as large as cropland, so this would become 1.86 (= 1 + 0.855) times as large given full replacement.
  - I am not claiming full replacement would be possible or needed, but the above illustrates there is great margin to increase cropland.
- "Scenarios assume that all stored food is consumed in Year 1", so there is room for better rationing.
- "We do not consider farm-management adaptations such as changes in cultivar selection, switching to more cold-tolerating crops or greenhouses31 and alternative food sources such as mushrooms, seaweed, methane single cell protein, insects32, hydrogen single cell protein33 and cellulosic sugar34".
- "Large-scale use of alternative foods, requiring little-to-no light to grow in a cold environment38, has not been considered but could be a lifesaving source of emergency food if such production systems were operational".
- "Byproducts of biofuel have been added to livestock feed and waste27. Therefore, we add only the calories from the final product of biofuel in our calculations". However, it would have been better to redirect to humans the crops used to produce biofuels.
- It is possible to have a relatively low famine death rate with a calorie consumption lower than 1.91 k kcal/person/d:
  - The calorie supply (to households) in the Central African Republic (CAR) in 2015 was 1.73 k kcal/person/d. I assume household waste is quite negligible there, such that the calorie consumption is similar to the calorie supply.
  - The deaths from protein-energy malnutrition there in that year were 1.38 k, equal to 0.0286 % (= 1.38*10^3/(4.82*10^6)) of CAR's population in 2015. For context, global deaths from protein-energy malnutrition in 2019 were 238 k, equal to 0.00307 % (= 238*10^3/(7.76*10^6)) of the global population.
  - One of the anonymous reviewers commented low reported calorie supply values like CAR's in 2015 are underestimates due to smuggling, which would imply a greater death rate from malnutrition than the above if the real supply matched the reported one. Yet, this

---

most people on the edge of starvation being subsistence farmers who are not much affected by market prices. Apparently, "roughly 65 percent of Africa's population relies on subsistence farming. Subsistence farming, or smallholder agriculture, is when one family grows only enough to feed themselves. Without much left for trade, the surplus is usually stored to last the family until the following harvest".

effect is offset by Xia 2022 not considering the underreported calories. In other words, it is still possible to have a relatively low famine death rate with a reported, if not actual, calorie consumption lower than 1.91 k kcal/person/d.

- The same reviewer commented that an actual calorie consumption of 1.7 k kcal/person/d "is not sustainable, and literally killed people in WW2", as described in Taste of War: World War II and the Battle for Food. I agree 1.7 k kcal/person/d is far from optimal for adults[15], but I doubt it would reduce life expectancy to less than 2 years, such that it could be sustained during the worst years of the nuclear winter in Xia 2022, 2 and 3. Calorie consumption in the coastal village of Kaul (Papua New Guinea) was 1.68 k kcal/person/d (= (1.94 + 1.42)/2) based on the mean values provided by Norgan 1974 for 51 adult men and 69 adult women[16].

Probability of human extinction given insufficient calorie production

I obtained a probability of human extinction given insufficient calorie production of 10^-6 (= 1/10^6), considering 1 M years is the typical lifespan of a mammal species[17]. For context:

- See Luisa Rodriguez' and Carl Shulman's general arguments and considerations about the possibility of civilisation collapse leading to extinction. Here are Luisa's:
  - "Historical survival and resilience".
  - "The grace period".
  - "With population loss comes "decorrelation" of survivors".
  - "Non-uniformity of the initial catastrophe's impacts".
  - "The population loss would have to be incredibly extreme to lead to extinction".
- Inequitable food distribution tendentially decrease extinction risk:

---

[15] From Akisaka 1996, "the energy intake of the Okinawan centenarians living at home was about 1,100 kcal/day for both sexes, which was similar to that of centenarians throughout Japan". I do not particularly trust this because food consumption was assessed based on self-reports. "The dietary survey was done by one 24h recall method, as was done for centenarians living throughout Japan (3)".

[16] In these studies, I am always worried about food consumption being estimated based on self-reports, but this should not be an issue in Norgan 1974. "All of the food eaten by each individual subject was weighed after cooking (where applicable) and immediately before consumption. Food consumed in the house was weighed on a robust Avery balance, weighing to 1 kg in 10 g divisions, using a large bowl scale-pan. The balances were frequently calibrated. Masses were recorded to the nearest 5 g. Left-over portions or inedible portions were also weighed and subtracted from the initial mass. Subjects were followed when they left the immediate vicinity of the house and food eaten away from the house was weighed on a portable Salter dietary balance weighing up to 500 g in 5 g divisions. A light plastic jug and plate were used for liquids such as coconut water".

There would also have been margin to further decrease calorie consumption via reducing physical activity. "The way of life for all the people was moderately active - more so in the highlands [not in Kaul] - since they were subsistence farmers cultivating their own gardens for food".

[17] Humans are a mammal species.

    - For example, with 1 k kcal/person/d and equitable distribution, everyone would starve because that is less than the resting energy expenditure, which is 1.14 k kcal/person/d according to Fig. 5 of Xia 2022.
    - Nonetheless, with inequitable distribution, there is room for part of the population to have enough calories. From Table S2 of the supplementary information of Xia 2022, Australia's major food crops and marine fish production in year 2 of a nuclear winter involving 47 and 150 Tg would be 36.0 % and 24.2 % higher than under normal conditions.
- My probability seems compatible with Luke Oman, one of the 3 authors of Robock 2007, having guessed a risk of human extinction of 0.001 % to 0.01 % for an injection of soot into the stratosphere of 150 Tg.
    - According to Fig. 5a of Xia 2022, 150 Tg would result in a calorie consumption 56.5 % (= 1.08*10^3/(1.91*10^3)) as large as that for 84.2 Tg given equitable food distribution, consumption of all edible livestock feed, and no household food waste.
    - So Luke's guess for the extinction risk would presumably be significantly lower for 84.2 Tg.

Grantmakers and donors interested in decreasing extinction risk had better focus on artificial intelligence instead of nuclear war

Supposedly cause neutral grantmakers aligned with effective altruism have influenced 15.3 M$[18] (= 0.03 + 5*10^-4 + 2.70 + 3.56 + 0.0488 + 0.087 + 5.98 + 2.88) aiming to decrease nuclear risk[19]:

- ACX Grants supported Morgan Rivers via a grant of 30 k$ in 2021 "to help ALLFED improve modeling of food security during global catastrophes" (the public write-up is 1 paragraph).
- Founders Pledge's Global Catastrophic Risks Fund advised on 2.70 M$ (= 0.2 + 2.50), supporting:
    - The Pacific Forum recommending a grant of 200 k$ in 2023 (1 sentence).
    - The Carnegie Endowment for International Peace recommending a grant of 2.50 M$ in 2024 (1 sentence).
- The Future of Life Institute (FLI) supported nuclear war research via 10 grants in 2022 totalling 3.56 M$ (1 paragraph each), of which 1 M$ was to support Alan Robock's and Brian Toon's research.
- The Long-Term Future Fund (LTFF) directed 48.8 k$ (= 3.6 + 5 + 40.2), supporting:

[18] Excluding the grants from Longview Philanthropy's Nuclear Weapons Policy Fund (NWPF), whose size is not publicly available, but I do not think including them would significantly change the total. The grants to decrease nuclear risk from Longview's Emerging Challenges Fund (ECF) only represent 0.569 % (= 0.087/15.3) of my total, and I guess NWPF has not granted more than 1 OOM more money than that linked to ECF's grants to decrease nuclear risk.

I included a grant of 500 $ made by the Effective Altruism Infrastructure Fund (EAIF), but decided not to describe it (besides mentioning the size here). This is quite small, so I was worried identifying the grantee could be a little mean. In addition, the grantee asked me not to mention the grant.

[19] I listed the grantmakers alphabetically.

    - ALLFED via a grant of 3.6 k$ in 2021 for "researching plans to allow humanity to meet nutritional needs after a nuclear war that limits conventional agriculture" (1 sentence).
    - Isabel Johnson via an "exploratory grant" of 5 k$ in 2022 for "preliminary research into the civilizational dangers of a contemporary nuclear strike" (1 sentence).
    - Will Aldred via a grant of 40.2 k$ in 2022 to "1) Carry out independent research into risks from nuclear weapons, [and] 2) Upskill in AI strategy" (1 sentence).
- Longview Philanthropy's Emerging Challenges Fund directed 87 k$ (= 15 + 52 + 20), supporting:
    - The Council on Strategic Risks via a grant of 15 k$ in 2022 (2 paragraphs).
    - The Carnegie Endowment for International Peace via a grant of 52 k$ in 2023 (3 paragraphs).
    - Decision Research via a grant of 20 k$ in 2023 (6 paragraphs).
- Longview Philanthropy's Nuclear Weapons Policy Fund has supported the Council on Strategic Risks, Nuclear Information Project, and Carnegie Endowment for International Peace (1 paragraph each).
    - For transparency, I encourage Longview to share on their website information about at least the date and size of the grants this fund made[20].
- Open Philanthropy has supported Alan Robock's and Brian Toon's research on nuclear winter via grants totalling 5.98 M$ (= 2.98 + 3), 2.98 M$ in 2017, and 3 M$ in 2020[21] (2 paragraphs each).
- The Survival and Flourishing Fund (SFF) has supported ALLFED via grants totalling 2.88 M$ (= 0.01 + 0.13 + 0.175 + 0.979 + 0.427 + 1.16), 10 k$ and 130 k$ in 2019, 175 k$ and 979 k$ in 2021, 427 k$ in 2022, and 1.16 M$ in 2023 (1 sentence each).

I encourage grantmakers to be more transparent by sharing further information about their grants. The extension of the public write-ups respecting the grants above ranged from 1 sentence to 6 paragraphs, with the median being 1 paragraph[22].

I consider the grant to Will was worth it, as I can see it having contributed to him now being a "researcher in longtermist AI strategy" at Metaculus. All of the others seem way less cost-effective than the current marginal grants of LTFF, which are overwhelmingly aimed at decreasing AI risk:

- I guess the nearterm annual extinction risk from AI is 1.69 M (= 10^-5/(5.93*10^-12)) times that from nuclear war. This assumes an nearterm annual extinction risk from AI of 0.001 %, which I motivate later in the section.
- I consider the annual spending on decreasing extinction risk from AI is 35.4 (= 4.04*10^9/(114*10^6)) times that on decreasing extinction risk from nuclear war. I determined this from the ratio between:

[20] I emailed Longview's Head of Grants Management & Compliance, Andrew Player, about this on 29 January 2024. He said it was a busy time, and that he would respond in due course.
[21] These grants were made in the context of Open Philanthropy's global catastrophic risks portfolio. In contrast, this and this grants to increase food resilience against abrupt sunlight reduction scenarios (nuclear, volcanic or impact winters) were made under the global health and wellbeing portfolio.
[22] 16th smallest/largest write-up of a total of 31. 12 were no longer than 1 sentence, and 26 no longer than 1 paragraph.

    - 4.04 G$ (4.04 billion USD) on nuclear risk in 2020, which I got from the mean of a lognormal distribution with 5th and 95th percentile equal to 1 and 10 G$, corresponding to the lower and upper bound guessed in 80,000 Hours' profile on nuclear war. "This issue is not as neglected as most other issues we prioritise. Current spending is between $1 billion and $10 billion per year (quality-adjusted)" (see details).
    - 114 M$ (= (79.8 + 32 + 2*1)*10^6) on "AI safety research that is focused on reducing risks from advanced AI" in 2023:
        - 79.8 M$ from the National Science Foundation (NSF), LTFF, Open Philanthropy, SFF and "other".
        - "~$32m per year" from "for-profit companies" (Anthropic, Conjecture, Google Deepmind and OpenAI).
        - 2 times "probably at least $1m per year" from "individual donors".
- So the nearterm annual extinction risk per annual spending for AI risk is 59.8 M (= 1.69*10^6*35.4) times that for nuclear risk.
- It would be super hard for the best interventions to decrease nuclear risk to be so many OOMs more tractable that they overturn the massive difference in importance and neglectedness illustrated above (relatedly).
- Consequently, I consider grantmakers and donors interested in decreasing extinction risk had better focus on AI instead of nuclear war.

Some caveats:

- I expect AI risk will become much less neglected in the next few decades, and the cost-effectiveness of interventions to decrease AI risk to significantly drop as a result.
- Interventions to decrease nuclear risk have indirect effects which will tend to make their cost-effectiveness more similar to that of the best interventions to decrease AI risk. I guess the best marginal grants to decrease AI risk are much less than 59.8 M times as cost-effective as those to decrease nuclear risk. At the same time:
    - I believe it would be a surprising and suspicious convergence if the best interventions to decrease nuclear risk based on the more direct effects of nuclear war also happened to be the best with respect to the more indirect effects. I would argue directly optimising the indirect effects tends to be better.
    - For example, I agree competition between the United States and China is a relevant risk factor for AI risk, and that avoiding nuclear war contributes towards a better relationship between these countries, thus also decreasing AI risk. Yet, in this case, I would expect it would be better to explicitly focus on interventions in AI governance and coordination, China-related AI safety and governance paths, understanding India and Russia better, and improving China-Western coordination on global catastrophic risks.
- It can still make sense for cause neutral grantmakers to recommend donors who are not so to support interventions to decrease nuclear risk[23]. The alternative may well be less cost-effective, and supporting interventions to decrease nuclear risk could be a pathway towards influencing more pressing areas.

[23] For example, Founders Pledge is a cause neutral organisation that advises some donors who are not so, such as ones partial to climate. Longview Philanthropy is a philanthropic advisory service, so I guess it operates under similar constraints, supporting some donors who are not cause neutral.

I arrived at a nearterm annual extinction risk from AI of 0.001 % as follows. I think looking into how species have gone extinct in the past is the best reference class to estimate AI risk. Jacob Steinhardt did an analysis which has some relevant insights:
"Thus, in general most species extinctions are caused by:

- A second species which the original species has not had a chance to adapt to. This second species must also not be reliant on the original species to propagate itself.
- A catastrophic natural disaster or climate event.
- Habitat destruction or ecosystem disruption caused by one of the two sources above."

I believe we have pretty good reasons to think the 2nd point applies much more weakly to humans than animals, but the 1st holds if one sees advanced AI as analogous to a new species[24]. I would still claim deaths in past terrorist attacks and wars provide a strong basis for arguing that humans will not go extinct via an AI war or terrorist attack. However, the 1st point alludes to what seems to me to be the greatest risk from AI, natural selection favouring AIs over humans. Since 1 M years is the typical lifespan of a mammal species, my prior extinction risk from AI in a random year this century is 10^-6 (= 1/10^6). Further accounting for inside view considerations, I guess the extinction risk from AI in a random year from 2025 to 2050 is 0.001 %. Relatedly, I encourage readers to check Zach Freitas-Groff's post on AGI Catastrophe and Takeover: Some Reference Class-Based Priors.

I should note I do not consider humans being outcompeted by AI as necessarily bad (relatedly). I strongly endorse expected total hedonistic utilitarianism (ETHU), and I would be surprised if humans were the most efficient way of increasing welfare longterm. At the same time, minimising nearterm extinction risk from AI seems like a good heuristic to align it with ETHU.

The case for sometimes prioritising nuclear extinction risk over AI extinction risk is much weaker than the case for sometimes prioritising natural extinction risk over nuclear extinction risk

Cost-effectiveness of decreasing extinction risk from nuclear war

I guess lobbying for nuclear arsenal limitation is one of the most cost-effective interventions to decrease nearterm extinction risk from nuclear war. The Centre for Exploratory Altruism Research (CEARCH) estimated it averts disability-adjusted life years (DALYs) 5.25 k times as cost-effectively as GiveWell's top charities, although:
"The headline cost-effectiveness will almost certainly fall if this cause area is subjected to deeper research: (a) this is empirically the case, from past experience; and (b) theoretically,

[24] I would update towards a higher extinction risk from wars relative to advanced AI systems if interspecific competition was more common relative to intraspecific one.

we suffer from optimizer's curse (where causes appear better than the mean partly because they are genuinely more cost-effective but also partly because of random error favouring them, and when deeper research fixes the latter, the estimated cost-effectiveness falls).”

Despite this, lobbying for nuclear arsenal limitation still looks promising among interventions to decrease nuclear risk. For context, CEARCH estimated, subject to the caveat above too, that conducting a pilot study of a resilient food source would be 14 times as cost-effective as GiveWell's top charities, i.e. just 0.267 % (= 14/(5.25*10^3)) as cost-effective as lobbying for nuclear arsenal limitation.

CEARCH determined lobbying for nuclear arsenal limitation decreases 9*10^-10 of the nuclear risk per dollar, but I guess the actual cost-effectiveness is only 1 % as high, such that it is only 52.5 (= 0.01*5.25*10^3) times as cost-effective as GiveWell's top charities at averting DALYs. Consequently, I guess lobbying for nuclear arsenal limitation decreases 9*10^-12 (= 0.01*9*10^-10) of the nuclear risk per dollar, which respects a cost-effectiveness of decreasing nearterm extinction risk from nuclear war of 5.34*10^-7 bp/T$[25] (= 9*10^-12*5.93*10^-12).

Cost-effectiveness of decreasing extinction risk from asteroids and comets

Salotti 2022 estimated the extinction risk from 2023 to 2122 from asteroids and comets is 2.2*10^-12 (see Table 1). This comes from the probability of long period comets with a diameter larger than 100 km colliding with Earth[26], for which the warning time is shorter than 5 years (see Table 1). The nearterm annual extinction risk from asteroids and comets respecting Salotti 2022 is 2.20*10^-14 (= 1 - (1 - 2.2*10^-12)^(1/100)).

Jean-Marc Salotti, the author of Salotti 2022, guesses it would cost hundreds of billions of dollars to design and test shelters which would decrease the extinction risk from asteroids and comets by 50 %[27]. I supposed a cost of 182 G$ (= 2/(1/10^3 + 1/100)*10^9), which is the reciprocal of the mean of the reciprocal of a uniform distribution ranging from 100 to 1 k

---

[25] 1 bp/T$ corresponds to 0.01 percentage points per 1 trillion dollars.

[26] One of the anonymous reviewers guessed comets larger than 10 km would still have a 20 % chance of causing extinction while being 500 times as likely as ones larger than 100 km. This would imply the extinction risk from comets larger than 10 km being 100 (= 0.2*500) times as large as that from ones larger than 100 km. As a result, the point I am making in this section would become stronger by 2 OOMs.

Salotti 2022 justifies the threshold of 100 km as follows:
“A 10 km sized asteroid could threaten large populations on Earth but there would still exist safe places on Earth to survive (Sloan et al., 2017, Toon et al., 1994, Chapman and Morrison, 1994, Mathias et al., 2017, RUMPF et al., 2017, Collins et al., 2005).”

I opted to rely on Salotti 2022's mainline estimate in my post, but I have not looked into the studies above. Less importantly, I also think a lower extinction risk per time makes more sense for shorter periods, given a less strict requirement for extended survival, and my nearterm annual extinction risk from nuclear war respects a period of 26 years (= 2050 - 2025 + 1), whereas Salotti 2022's estimate concerns one of 100 years.

[27] Information provided via email.

G$[28]. So I guess the cost-effectiveness of decreasing nearterm extinction risk from asteroids and comets is 6.04*10^-10 bp/T$ (= 0.50*2.20*10^-14/(182*10^9)).

Comparisons

According to my estimates, the cost-effectiveness of decreasing nearterm extinction risk from nuclear war via lobbying for nuclear arsenal limitation is 884 (= 5.34*10^-7/(6.04*10^-10)) times that from decreasing nearterm extinction risk from asteroids and comets via shelters. However, I do not think one can conclude from this high ratio that lobbying for nuclear arsenal limitation is better than working on shelters, as these would decrease extinction risk from not only asteroids and comets, but also other risks, including nuclear war.

On the other hand, I would say the case for sometimes prioritising nuclear extinction risk over AI extinction risk is much weaker than the case for sometimes prioritising natural extinction risk over nuclear extinction risk:

- The ratio of 884 between the cost-effectiveness of decreasing nuclear and asteroids and comets risk is many orders of magnitude lower than the ratio of 59.8 M I calculated between the nearterm annual extinction risk per annual spending of AI and nuclear risk.
- The conclusion just above is reinforced if one believes there are more pressing natural risks besides those from asteroids and comets. According to Toby Ord's guesses given in The Precipice, the existential risk from 2021 to 2120 from supervolcanic eruptions, his largest natural risk, is 100 (= 10^-4/10^-6) times that from asteroids and comets.
    - However, I am not that moved by Toby's estimate for the existential risk from supervolcanic eruptions.
    - I believe extinction risk from these is many OOMs lower, as arguably proved to be the case for asteroids and comets.

Further research to increase the resilience of my cost-effectiveness estimates would be useful.

Extinction risk from nuclear war was massively overestimated in The Existential Risk Persuasion Tournament

I collected in the table below the predictions of the superforecasters, domain experts, general existential risk experts, and non-domain experts of XPT for the risk of human extinction from nuclear war. The estimates respect the medians across 88 superforecasters, 13 domain experts, 14 general existential risk experts, and 58 non-domain experts.

[28] I used this because E("cost-effectiveness") = E("benefits"/"cost") = E("benefits")*E(1/"cost") = E("benefits")/(1/E(1/"cost")), assuming benefits and cost are independent.

Period from 2023 to…	Total extinction risk from nuclear war[29]		Annual extinction risk from nuclear war[30]	
	Superforecasters	Domain experts	Superforecasters	Domain experts
2030	0.001 %	0.02 %	1.25*10^-6	2.50*10^-5
2050	0.01 %	0.12 %	3.57*10^-6	4.29*10^-5
2100	0.074 %	0.55 %	9.49*10^-6	7.07*10^-5
Period from 2023 to…	**Total extinction risk from nuclear war**		**Annual extinction risk from nuclear war**	
	General existential risk experts	**Non-domain experts**	**General existential risk experts**	**Non-domain experts**
2030	0.03 %	0.01 %	3.75*10^-5	1.25*10^-5
2050	0.17 %	0.07 %	6.08*10^-5	2.50*10^-5
2100	0.7 %	0.19 %	9.01*10^-5	2.44*10^-5

The superforecasters', domain experts', general existential risk experts', and non-domain experts' annual risk of human extinction from nuclear war from 2023 to 2050 is 602 k (= 3.57*10^-6/(5.93*10^-12)), 7.23 M (= 4.29*10^-5/(5.93*10^-12)), 10.3 M (= 6.08*10^-5/(5.93*10^-12)) and 4.22 M (= 2.50*10^-5/(5.93*10^-12)) times my nearterm annual risk. So I get a sense the extinction risk from nuclear war was massively overestimated in XPT. Do you agree? If yes, should one put little trust in other estimates of extinction risk from XPT? I think so. Still, I believe the XPT was quite valuable given the wealth of information shared in the report explaining the rationale for the forecasts (see Appendix 7).

One could argue the large gap between XPT's estimates and mine points to me not having sufficiently updated my independent impression. I agree epistemic deference is valuable in general, but it is unclear to me whether I should be deferring more:

- I am familiar with what informed XPT's nuclear extinction risk predictions, having read the respective sections "Sources of agreement, disagreement and uncertainty", "Arguments given for low-end forecasts", and "Arguments given for higher-end forecasts" (pp. 298 to 303).
- Some participants in the XPT seemed to believe in a much lower nuclear extinction risk than the medians I presented (emphasis mine):
    - "Most forecasters whose probabilities were near the median factored in a range of possible risks, including world wars, nuclear winters, and even artificial-intelligence-driven NERs [nuclear extinction risks], but concluded that **even under worst case scenarios, the extinction of humanity (give or**

[29] See pp. 293 and 294.
[30] "Annual risk" = 1 - (1 - "total risk")^(1/"duration of the period in years"). The periods have durations of 8 (= 2030 - 2023 + 1), 28 (= 2050 - 2023 + 1) and 78 (= 2100 - 2023 + 1) years.

take 5000 people) would be near impossible...even if an NER [nuclear existential risk] had set humanity on a path that made eventual extinction a foregone conclusion, existing resources on earth would allow at least 5000 survivors to hang on for seventy-eight years".
    - "For many, the thought of getting to **less than 5000 humans alive** was **simply too far fetched** an outcome and they couldn't be persuaded otherwise in what they saw as credible scenarios".
    - "[T]he set of circumstances required for this to happen are quite low, though obviously not impossible. These circumstances are that there will be a nuclear conflict between 2 nations both capable and willing to fire at everyone everywhere between the two of them: 'very bad case scenarios' where India and Pakistan, or the US and Russia, or China and anyone else, fired everything they had at just each other, or even at each other and each other's close allies, would likely not cause extinction…**it requires some of the big nuclear powers to decide to try to take literally everyone down with them, and that they actually succeed**".
    - "So we think that the probabilities in this question are dominated by scenarios of total nuclear war before 2050 which cause civilizational and climate collapse to the point where long-term survival becomes impossible to save for very well-prepared shelters. But even pessimistic scenarios seem unlikely to lead to a collapse that is fast enough to reduce the global population to below 5000 by 2100".
    - "**There aren't compelling arguments on the higher end** for this question again due to the fact that this is a very high bar to achieve".
    - "The team predicts that there will be pockets of people who survive in various regions of the world. Their survival may be at Neolithic standards, but there will be tribes of people who band together and restart mankind. After all, **many mammals survived the asteroid and ice age that killed the dinosaurs**".
    - "[A] certain number of team members feel that **even if there was a full strategic exchange and usage of all of the world's nuclear arsenal still humanity would be able to keep its numbers over 5000**. The argument for this is the number [a]nd population of uncontacted tribes, or isolated human populations like the Easter island population pre-contact, that have managed to hold numbers of over 5000 in extremely harsh conditions".
    - "**[A]lmost certainly some people would survive** on islands or in caves given even the worst of worst cases".
    - "Southern Hemisphere likely to be less impacted – New Zealand, Madagascar, Pacific Islands, Highlands of Papua New Guinea, unlikely to be targeted and include areas with little global and technology dependence…Just the population of Antarctica in its summer is ~5000 people. **Even small islands surviving could easily mean more than 5k people**".
    - "**[There are s]everal regions in the world that would not be affected by nuclear conflict directly and have decent climatic conditions** to support 100 of millions even in a NW [nuclear winter]".
- I believe my estimate involved much more explicit modelling than XPT's.

- There is very little formal evidence on the accuracy of forecasting very rare events like human extinction[31].
- In general, I suspect there is a tendency to give probabilities between 1 % and 99 % for events whose mechanics we do not understand well, like the factors involved in a product to estimate the chance of extinction.
    - Such a range encompasses the vast majority (98 % = 0.99 - 0.01) of the available linear space (from 0 to 1), and forecasting questions are often formulated with the aim of reasonable predictions falling in that range.
    - However, the available logarithmic space is infinitely vast, and it is hard to rule out an astronomically low extinction risk. In contrast, extinction risk could be overly high if it implies a too low probability of our current existence.
    - So there is margin for moderate guesses (e.g. between 1 % and 99 %) to be major overestimates.

As a side note, the superforecasters predicted the annual risk from 2023 to 2100 is 7.59 (= 9.49*10^-6/(1.25*10^-6)) times that from 2023 to 2030, the domain experts 2.83 (= 7.07*10^-5/(2.50*10^-5)) times, the general existential risk experts 2.40 (= 9.01*10^-5/(3.75*10^-5)) times, and the non-domain experts 1.95 (= 2.44*10^-5/(1.25*10^-5)) times, i.e. all expected the risk to increase throughout this century. Interestingly, none foresaw major changes to the median number of nuclear warheads by 2040, which is some evidence against large increases in nuclear arsenals. Relative to the 12,705 in 2022 (see pp. 532 and 533):

- 31 superforecasters predicted 13,500, i.e. an increase of 6.26 % (= 13,500/12,705 - 1).
- 1 domain expert predicted 11,990, i.e. a decrease of 5.63 % (= 1 - 11,990/12,705).
- 5 general existential risk experts predicted 10,200, i.e. a decrease of 19.7 % (= 1 - 10,200/12,705).
- 10 non-domain experts predicted 12,952.5, i.e. a decrease of 1.95 % (= 1 - 12,952.5/12,705).

Consequently, I think the superforecasters, domain experts, general existential risk experts, and non-domain experts implicitly predicted at least one of the following. Nuclear war becoming more frequent, having a greater potential to escalate[32], or humanity becoming less resilient to it. I only seem to agree with the 2nd of these.

Toby Ord greatly overestimated tail risk in The Precipice

I collected in the table below Toby's annual existential risk from 2021 to 2120 from AI, nuclear war, and asteroids and comets based on his guesses given in The Precipice. I also added my estimates for the nearterm annual extinction risk from the same 3 risks, and the ratio between Toby's values and mine. The values are not directly comparable, because

[31] Additionally, there is very little formal evidence on the accuracy of long-range forecasting (I am only aware of Tetlock 2023), but this is arguably not as important because I am only relying on XPT's extinction risk until 2030.

[32] Including via a more heavy-tailed distribution of the number of nuclear warheads.

Toby's refer to existential risk and mine to extinction risk. Nonetheless, I still have the impression Toby greatly overestimated tail risk. This is in agreement with David Thorstad's series exaggerating the risks, which includes subseries on climate, AI and bio risk, and discusses Toby's book The Precipice.

Risk[33]	Toby's annual existential risk from 2021 to 2120[34]	My nearterm annual extinction risk	Ratio between Toby's value and mine
AI	0.105 %	0.001 %	105
Nuclear war	1.00*10^-5	5.93*10^-12	1.69 M
Asteroids and comets	1.00*10^-8	2.20*10^-14	455 k

The estimates of the tail risk from asteroids and comets are arguably the most robust, so it is interesting there is a large difference between Toby's and mine even there. There are many concepts of existential catastrophe[35], but I do not think one can say existential risk from asteroids and comets is anything close to 455 k times as high as extinction risk from these:

- In The Precipice, Toby says the probability of an asteroid larger than 10 km colliding with Earth in the next 100 years is lower than 1 in 150 M (Table 3.1), and guesses that the risk from comets larger than 10 km is similarly large (p. 72), which implies a total risk from asteroids and comets larger than 10 km of around 1.33*10^-8 (= 2/(150*10^6)). This is only 1.33 % (= 1.33*10^-8/10^-6) of Toby's guess for the existential risk from asteroids and comets, which implies Toby expects the vast majority of existential risk to come from asteroids and comets smaller than 10 km.
- The last mass extinction "was caused by the impact of a massive asteroid 10 to 15 km (6 to 9 mi) wide", and happened 66 M years ago. It involved an impact winter, which played a role in the extinction of the dinosaurs, and may well have contributed to the emergence of mammals and ultimately humans.
- So Toby would expect an asteroid impact similar to that of the last mass extinction to be an existential catastrophe. Yet, at least ignoring anthropics, I believe the probability of not fully recovering would only be 0.0513 % (= e^(-10^9/(132*10^6))), assuming:
    - An exponential distribution with a mean of 132 M years (= 66*10^6*2) represents the time to go from i) human extinction due to such an asteroid to ii) evolving a species as capable as humans at steering the future. I supposed this on the basis that:
        - An exponential distribution with a mean of 66 M years describes the time between extinction threats as well as that to go from i) to ii) conditional on no extinction threats.
        - Given the above, extinction and full recovery are equally likely. So there is a 50 % chance of full recovery, and one should expect the

[33] Ordered from the largest to the smallest.
[34] "Annual risk" = 1 - (1 - "total risk")^(1/"duration of the period in years"). The period has a duration of 100 years (= 2120 - 2021 + 1).
[35] I prefer focussing on clearer metrics.

time until full recovery to be 2 times (= 1/0.50) as long as that conditional on no extinction threats.
    - The above evolution could take place in the next 1 billion years during which the Earth will remain habitable.
- In addition, one should arguably suppose a species as capable as humans at steering the future would have similarly good values, even if different.
- Setting the existential risk from asteroids and comets to the extinction risk estimated in Salotti 2022 seems much more legitimate, as it relies on a threshold of 100 km for the impactor. This is 1 OOM larger, and 3 OOMs more energetic[36] than the asteroid involved in the last mass extinction, thus having the potential to cause the extinction of not only humans, but also of many other species in our evolutionary path.

Interventions to decrease deaths from nuclear war should be assessed based on standard cost-benefit analysis

I believe interventions to decrease deaths from nuclear war should be assessed based on standard cost-benefit analysis (CBA):

- Having in mind my astronomically low nearterm annual extinction risk from nuclear war, it is unclear to me whether interventions to decrease deaths from nuclear war decrease extinction risk more cost-effectively than broader ones, like the best interventions to boost economic growth or decrease disease burden (e.g. GiveWell's top charities).
- I expect extinction risk can be decreased much more cost-effectively by focussing on AI risk rather than nuclear risk. So I would argue interventions to decrease deaths from nuclear war can only be competitive under an alternative worldview, like ones where the goal is boosting economic growth or decreasing disease burden.

Moreover, I would propose using standard CBAs not only in the political sphere, as argued by Elliott Thornley and Carl Shulman, but also outside of it. In terms of what grantmakers aligned with effective altruism have been doing[37]:

- CEARCH has done standard CBAs:
    - Shallow Report on Nuclear War (Abolishment) by Joel Tan (the cost-effectiveness was estimated to be 0.4 times that of GiveWell's top charities).
    - Shallow Report on Nuclear War (Arsenal Limitation) by Joel Tan (5 k times that of GiveWell's top charities).
    - Intermediate Report on Abrupt Sunlight Reduction Scenarios by Stan Pinsent (14 times that of GiveWell's top charities).
- Founders Pledge has done a standard CBA:

[36] Kinetic energy is proportional to mass, and the mass of a sphere is proportional to its diameter to the power of 3. Kinetic energy is also proportional to speed to the power of 2, but I am guessing the impact speed is independent of the size.
[37] I listed the grantmakers alphabetically.

    - Doubling risk reduction spending (2.5 times that of Against Malaria Foundation).
- Open Philanthropy has made grants:
    - In the area of scientific research, under their global health and wellbeing portfolio, which tends to rely on standard CBA:
        - Penn State University — Emergency Food Research (Charles Anderson) (109 k$).
        - Penn State University — Research on Emergency Food Resilience (Charles Anderson) (2020) (3.06 M$).
    - Under their global catastrophic risks portfolio, which does not tend to rely on standard CBA:
        - Rutgers University — Nuclear Conflict Climate Modeling (2.98 M$).
        - Rutgers University — Nuclear Conflict Climate Modeling (2020) (3 M$).

I wonder whether the best interventions to decrease deaths from nuclear war would, based on in-depth CBAs, be better than donating to GiveWell's All Grants Fund. From the ones above, I guess only nuclear arsenal limitation would be so.

Increasing calorie production via new food sectors is less cost-effective to save lives than measures targeting distribution

Nuclear winter is a major source of risk of global catastrophic food failures. Nonetheless, my estimates imply the annual probability of a nuclear war causing (globally) insufficient calorie production is 0.0553 % (= 0.0131*0.0422). This suggests food distribution rather than production will be the bottleneck to decrease famine deaths in the vast majority of circumstances, as is the case today[38]. So I think increasing calorie production via new (or massively scaled up) food sectors, like greenhouse crop production, lignocellulosic sugar, methane single cell protein or seaweed, is less cost-effective to save lives than measures targeting distribution, like ones aiming to ensure the continuation of international food trade.

One of the anonymous reviewers commented the aforementioned new food sectors "are definitely helpful for loss of international trade scenarios". I suspect the reviewer has something like the following in mind:

- From Fig. 5b of Xia 2022, the minimum soot injected into the stratosphere for insufficient calorie consumption is 10 Tg[39] given no international food trade, consumption of all edible livestock feed, and no household food waste.
- In contrast, I estimated a minimum soot injected into the stratosphere for insufficient calorie production of 84.2 Tg, supposing the net effect on calorie production of all the adaptation measures is similar to assuming equitable food distribution, consumption of all edible livestock feed, and no household food waste.

[38] Note I am not arguing prices will stay constant. I am claiming prices will go up mostly due to limitations in food distribution rather than production.
[39] Eyeballed.

I think the reviewer may be concluding from the above that, given no international food trade, calorie consumption would be much lower, and therefore increasing food production via new food sectors would become much more important relative to distribution. I agree with the former, but not the latter. Loss of international food trade is more of a problem of food distribution than production. If this increased thanks to new food sectors, but could not be distributed to low-income food-deficit countries (LIFDCs) due to loss of trade, there would still be many famine deaths there. Many LIFDCs are in tropical regions too, where there is a smaller decrease in crop yields during a nuclear winter (see Fig. 4 of Xia 2022).

Furthermore, greater loss of trade and supply chain disruptions will be associated with greater loss of population and infrastructure, which in turn will arguably make solutions relying on new food sectors less likely to be successful relative to ones leveraging existing sectors. Examples of the latter include decreasing animal and biofuel production which relies on edible crops, expanding crop area, and using more cold-tolerant crops.

My point about distribution rather than production being a bottleneck loses strength as the severity of the nuclear winter increases. For an injection of soot into the stratosphere of 150 Tg, the calorie consumption given equitable food distribution, consumption of all edible livestock feed, and no household food waste would be 1.08 k kcal/person/d (see Fig. 5a of Xia 2022), which is just 56.5 % (= 1.08*10^3/(1.91*10^3)) of the minimum caloric requirement. Producing more calories would be crucial in this case. Moreover, Xia 2022's 150 Tg scenario involves 4.4 k nuclear detonations (see Table 1). The disruptions to international food trade caused by these would be so extensive that it would be especially useful for countries to have local resilience, such as by producing their own food.

Finally, there is a risk that focussing on new food sectors counterfactually increases the suffering of farmed animals without decreasing starvation (not to mention the meat-eater problem). Some countries may not need to consume all edible livestock feed to mitigate starvation, in which case increasing production from new food sectors could allow for greater consumption of farmed animals with bad lives. Somewhat relatedly, I have very mixed feelings about promoting resilient food solutions which rely on increasing factory-farming, such as ALLFED mentioning insects.

Acknowledgements

Thanks to Anonymous Person 1, Anonymous Person 2, Anonymous Person 3, Anonymous Person 4, Carl Robichaud, Ezra Karger, Farrah Dingal, Matthew Gentzel, Nuño Sempere and Ross Tieman for feedback on the draft[40]. Thanks to Jean-Marc Salotti for guessing the cost of shelters which would decrease the extinction risk from asteroids and comets.

[40] Names ordered alphabetically.