Founders Pledge

Global Catastrophic Nuclear Risk

A Guide for Philanthropists

CHRISTIAN RUHL

JULY 2023

Acknowledgements and Disclaimers

I am thankful to the many experts who have contributed to this project with their insight, feedback, and critiques, including in semi-structured interviews. **The analysis and views expressed in this report do not necessarily reflect the views of anyone consulted for the project.**

With special thanks to:

- Dr. James Scouras for his thoughtful advice and key insights on nuclear war as a global catastrophic risk and for serving as an external reviewer of the report.
- Matthew Gentzel for reviewing various drafts of the report, providing valuable feedback throughout the project, and serving as an external reviewer of the report.
- Dr. James Acton, Conor Barnes, Tom Barnes, Patty-Jane Geller, Matt Lerner, Dr. Jeffrey Lewis, Ankit Panda, Dr. Andrew Reddie, and Carl Robichaud for reviewing earlier versions of the "Right of Boom" analysis in this document and for their helpful comments and suggestions.
- Dr. Michael Horowitz for useful conversations and insights on these issues.
- Dr. Michael Cassidy and Lara Mani for their explanation of volcanic aerosol chemistry and the potentially self-limiting effects of some climate phenomena, and Dr. Jon Reisner, Professor Alan Robock and Professor Brian Toon for their discussions on nuclear winter.
- Dr. Jaime Yassif for many conversations on global catastrophic biological risks and general risk-reduction strategies.
- Matt Lerner for extensive feedback, edits and guidance on the report.
- Amber Dawn Ace for providing copy edits and helpful suggestions.
- Dr. Zoë Ruhl for her support and invaluable insight.

Parts of this document were previously published as a Founders Pledge report on "Philanthropy to the Right of Boom," which can be found here.

This report contains discussions of war and human suffering.

Executive Summary

Nuclear weapons are a global catastrophic risk; a nuclear war could kill untold millions, inflict horrific suffering on survivors, and derail human civilization as we know it. This report forms a guide for philanthropists who seek to mitigate this risk and maximize the counterfactual impact of their charitable donations. Specifically, the report seeks to guide funders entering this field in the wake of several challenges: the apparent collapse of post-Cold War arms control, the second year of the Russo-Ukrainian War, rising U.S.-China tensions, and a large funding shortfall for nuclear security. It mirrors many of the themes of Founders Pledge's *Guide to the Changing Landscape of High-Impact Climate Philanthropy,* and is indebted to the insights in that document.[1]

The report's analysis has four steps:[2]

1. Understanding key features of the landscape of nuclear philanthropy, with special attention to recent funding shortfalls.
2. Analyzing the structure of the problem, emphasizing the super-linearity of expected costs; not all nuclear wars are equal, and bigger nuclear wars could be disproportionately more damaging than smaller nuclear wars for both current generations and the long-term future.
3. Sketching guiding principles for nuclear philanthropy based on these ideas:
    - Prioritize minimizing expected global war damage;
    - Prioritize neglected strategies;
    - Multiply impact by shaping great power behavior;
    - Exercise leverage via policy advocacy;
    - Pursue a strategy of "robust diversification."
4. Exploring practical implications of these principles. The section briefly describes concrete projects that philanthropists can support. A conclusion enumerates key uncertainties and sketches a path forward for philanthropists.

Overall, the report argues that deep uncertainty surrounds nuclear risk, that shaky assumptions underpin much of the conventional wisdom on nuclear war, and that effective philanthropists must learn to leverage their donations despite this uncertainty. Although the report reflects the input of a variety of experts, it is only one approach to the problem. We hope and expect to revise its conclusions as we encounter new evidence.

[1] Johannes Ackva, Luisa Sandkühler, and Violet Buxton-Walsh, "A Guide to the Changing Landscape of High-Impact Climate Philanthropy" (Founders Pledge, November 2021).
[2] Technical appendices include calculations and further discussion of key issues relevant to nuclear risk reduction.

External Reviews

Founders Pledge's research reports undergo several rounds of internal and external review. To provide the reader context for this report, we have asked two outside experts to briefly write up their impressions:

Reviewer 1: James Scouras

James Scouras *is a senior scholar at the Johns Hopkins University Applied Physics Laboratory and the former chief scientist of the Defense Threat Reduction Agency's Advanced Systems and Concepts Office. Previously, he was program director for risk analysis at the Homeland Security Institute, held research positions at the Institute for Defense Analyses and the RAND Corporation, and lectured on nuclear policy in the University of Maryland's General Honors Program. Among his publications are the book* A New Nuclear Century: Strategic Stability and Arms Control*, coauthored with Stephen Cimbala, and the edited volume* On Assessing the Risk of Nuclear War*. Dr. Scouras earned his PhD in physics from the University of Maryland.*[3]

Scouras Review

Among the many and varied global catastrophic risks faced by humanity, nuclear war stands out for the combination of its potential immediacy, the horrific nature of its consequences, its long-term threat to civilization, and — most important of all — the fact that it has been created by humans and is subject to human interventions. With the end of the Cold War and the emerging multipolar nuclear future, a renaissance of creative, disciplined thinking is urgently needed and needs to be underwritten by both governmental and philanthropic organizations. Christian Ruhl's *Global Catastrophic Nuclear Risk: A Guide for Philanthropists* lays a thoughtful intellectual foundation for justifying and focusing impactful philanthropy on reducing nuclear risks.

Most important, this paper challenges conventional wisdom in significant ways. In particular, it recommends emphasizing "right-of-boom" thinking. For too long the United States and its allies have put all their eggs in the deterrence basket. If the success of deterrence could be guaranteed, there would be no need to worry about the aftermath of its failure. But the complacency over the robustness of deterrence that has emerged from over three-fourths of a century without nuclear war may not serve us well in the future. Ominous trends in horizontal and vertical proliferation, the intensification and periodic eruption of enduring interstate disputes, and the never-ending advent of potentially destabilizing technologies all suggest we need to be better prepared for the possibility of

[3] This biography is from *Nuclear War as a Global Catastrophic Risk*, modified in consultation with James Scouras.

nuclear war. Thus, we need to focus more on preventing small nuclear wars from escalating to large ones and recovering from all levels of nuclear war.

In addition, the paper is also innovative in its recommended focus on large nuclear wars that are disproportionately harmful compared to smaller nuclear wars. Large states can endure a small nuclear war, horrific though it will be. But it's improbable that nuclear combatant states could survive a war that unleashed the arsenals of the major nuclear powers, and it is not clear how many centuries civilization would be thrown back and for how long. Thus, avoiding and recovering from large nuclear wars needs much greater focus in nuclear policy.

Finally, the paper identifies what is the most challenging problem in nuclear strategy: how to maintain deterrence and stability in the emerging tripolar nuclear world, with the United States, Russia, and China possessing comparably large nuclear arsenals. The fundamental problem is that the forces that underwrite deterrence against any single nuclear state would be inadequate to deter a coordinated attack by two peer adversaries. This concern could spark an arms race and/or motivate destabilizing force postures, launch decisions, and targeting doctrines. While the United States Strategic Command and policy organs of the US Department of Defense have recognized this challenge, a workable approach has not been identified, and any consensus is not on the horizon.

All these issues and many others identified in Ruhl's paper require the sustained attention of the most knowledgeable, the most disciplined, and the most creative minds. Philanthropy can serve the critical roles of encouraging such minds to focus on this critical problem, fostering unconventional thinking, and challenging unimaginative, even wrongheaded and dangerous, government policies. Nuclear risk is far too important to leave to the generals.

Reviewer 2: Matthew Gentzel

***Matthew Gentzel** is a nuclear weapons policy program officer at Longview Philanthropy, where he co-leads Longview's work on nuclear issues. His prior work spanned emerging technology threat and policy assessment, with a particular focus on how advancements in AI may shape the future of influence operations, nuclear strategy, and cyber attacks. He has worked as a policy researcher with OpenAI, as an analyst in the US Department of Defense's Innovation Steering Group, and as a director of research and analysis at the US National Security Commission on Artificial Intelligence. Matthew holds an MA in strategic studies and international economics from Johns Hopkins SAIS and a BS in fire protection engineering from the University of Maryland.*[4]

Gentzel Review

This guide is one of the most comprehensive approaches I've seen to reducing global catastrophic nuclear risk from a "big picture" perspective. While no document can cover every angle, it brings a variety of strategic considerations and thought tools together with sufficient depth to help philanthropists develop effective nuclear risk reduction strategies.

The extreme non-linearity of nuclear scenarios, the impact of different nuclear risk reduction interventions on other catastrophic threats such as bioweapons, and prioritization under extreme uncertainty are a few among many themes the guide covers. With examples such as the Cooperative Threat Reduction program, the guide also illustrates how philanthropists can drive extremely cost-effective risk reduction measures by bringing attention to neglected issues early and leveraging government resources.

One idea that may deserve future attention is how cultivating talent with broad awareness of crucial considerations can help uncover further opportunities both to prevent and to reduce the damage of nuclear war. While the prevention of nuclear use is relatively not neglected within the field, people with *neglected expertise* in positions of influence may still find traction. Similarly, talented individuals can help sort through interventions that initially appear to have an ambiguous track record: identifying the conditions that determine success with high quality analysis. Opportunities of these sorts often lie hidden by a variety of factors: ideological polarization, analytic siloing, misaligned bureaucratic incentives, efforts to manipulate risk perception, and the constraints of state secrecy. Training and elevating experts that can sort through these factors and the other considerations put forward in this guide could be a great investment.

---

[4] This biography is from Longview Philanthropy's website, modified in consultation with Matthew Gentzel.

Introduction: Deep Uncertainty

Key Points:

- Deep uncertainty surrounds both the probability and the consequences of nuclear war.
- This uncertainty extends to the effectiveness of different risk-reduction measures that philanthropists could pursue.
- Analyzing potential "impact multipliers" derived from the problem's structure can guide effective philanthropy amidst this uncertainty.

We know very little about nuclear war.[5] Despite countless books, articles, game-theoretic models, war games, and billions of dollars spent on understanding risk reduction, even well-respected experts and seasoned policymakers face fundamental and often unresolvable uncertainties about nuclear war.[6]

The world has had only one experience with the horrors of nuclear war through the atomic bombings of Japan in the summer of 1945. This distinguishes nuclear risk from some other global catastrophic risks, such as biological risks, where the history of natural pandemics can help analysts anchor risk estimates, including estimates of risk from unprecedented engineered pandemics. Moreover, uncertainty on nuclear war arises not just from complexity, but from sometimes being an optimization target for states; nuclear postures and policies are fundamentally about shaping risk perception, which complicates accurate risk estimation.[7] These factors lead to three fundamental uncertainties:

1. Uncertainty on the probabilities of nuclear wars[8]

---

[5] "We" in this case refers to everyone concerned with nuclear risk reduction: philanthropists, global priorities researchers, decision-makers, policy advocates, think tank analysts, military officials, and the public at large.
[6] To illustrate this problem, Herman Kahn, the Cold War strategist and author of foundational texts in the field such as *On Thermonuclear War,* liked to ask, "How many thermonuclear wars have you fought recently?" Similarly, the RAND scholar and Department of Defense policymaker Alain Enthoven reportedly once snapped, "General, I have fought just as many nuclear wars as you have." (Sharon Ghamari-Tabrizi, "Simulating the Unthinkable: Gaming Future War in the 1950s and 1960s," *Social Studies of Science* 30, no. 2 (2000): 165; Matthew Connelly et al., "'General, I Have Fought Just as Many Nuclear Wars as You Have': Forecasts, Future Scenarios, and the Politics of Armageddon," *The American Historical Review* 117, no. 5 (2012): 1431–60.)
[7] Thanks to Matthew Gentzel for this point.
[8] This report refers to "probabilities" and "wars" in the plural because different kinds of nuclear wars can have qualitatively different consequences for humanity, as explained below in "Nuclear Wars Are Not Created Equal."

2. Uncertainty on the consequences of nuclear wars
3. Uncertainty on effective risk-reduction measures

Together, these factors make up the fundamental components of risk (as a function of probability and consequence) and risk reduction. Such uncertainty can inspire apathy about risk reduction — in Cold War America, for example, popular jokes mocked attempts to reduce the consequences of nuclear war as futile: "What do you do when you see the flash? You put your head between your legs and kiss your ass goodbye."[9] We need not take this fatalistic view. As explained throughout this report, philanthropists and policymakers can still prioritize interventions by understanding the general structure of the problem.

The following three sections disaggregate (1) uncertainty on probabilities, (2) uncertainty on consequences, and (3) uncertainty on risk reduction.

Uncertainty on Probabilities

First, how likely is nuclear war? We do not really know.[10] Ways of assessing this probability include:

- Crowdsourced probabilistic forecasting;[11]
- Elicited expert knowledge;[12]
- Naïve base-rate forecasting;[13]
- Subjective policymaker judgments;[14]
- Probabilistic risk assessment based on near-misses and "teetering coin" analyses;[15]

---

[9] Quoted in Edward Geist, *Armageddon Insurance: Civil Defense in the United States and Soviet Union, 1945–1991*, 2019, 11.

[10] For a more detailed treatment of this subject, see James Scouras, ed., *On Assessing the Risk of Nuclear War* (Johns Hopkins Applied Physics Laboratory.).

[11] See, e.g. "Ragnarök Question Series: If a Nuclear Catastrophe Occurs, Will It Reduce the Human Population by 95% or More?," Metaculus, November 22, 2018, https://www.metaculus.com/questions/1585/ragnar%25C3%25B6k-question-series-if-a-nuclear-catastrophe-occurs-will-it-reduce-the-human-population-by-95-or-more/. As of May 2023, the prediction stood at 5%.

[12] See Jane M. Booker, "Elicited Expert Knowledge" in *On Assessing the Risk of Nuclear War.*

[13] That is, there has been one nuclear war between 1945 and 2023, leading to a very simplistic estimate of 1-in-77.

[14] Like Kennedy's famous "between 1-in-3 and even" statement. For more, see Scouras, "Framing the Questions" in *On Assessing the Risk of Nuclear War.*

[15] "[It] is possible to reclaim valuable information by looking not only at whether each toss showed heads or tails but also at the nuances of how the coin behaved during that toss. If all sixty-six tosses immediately landed on tails without any hesitation, that would be evidence that the coin was more strongly weighted in favor of tails and, thus, additional evidence that Schlesinger was right. Conversely, if any of the tosses teetered on the coin's edge, leaning first one way and then the other before finally showing tails, that would be evidence in favor of McNamara's position [that perpetual reliance on nuclear deterrence will "destroy nations"]." Martin E. Hellman, "Probabilistic Risk Assessment," in *On Assessing the Risk of Nuclear War,* 88.

- and more, including the aggregation of multiple methods.

Aggregation of multiple forecasts can yield rough point estimates. Global priorities researcher Luisa Rodriguez, for example, has aggregated several estimates with the arithmetic mean of probabilities for an annualized probability of 1.1% of nuclear war, and a 0.38% probability of U.S.-Russia war.[16] A spreadsheet in the appendix adds further estimates to Rodriguez's table and aggregates the estimates using the geometric mean of odds rather than the arithmetic mean of probabilities, yielding a 0.986% annual probability of nuclear war.[17] Point estimates can, however, obscure the magnitude of the uncertainty that surrounds these questions. Probability ranges can provide a more complete picture; Martin Hellman has given a rough order-of-magnitude estimate of around 1% per year in a publication by the Johns Hopkins Applied Physics Laboratory (APL), with a lower bound of 0.1% per year and an upper bound of 10% per year, based on reasoning about past crises and civilization's survival with nuclear weapons so far.[18]

These estimates (including the estimates in this report) remain deeply flawed. There may be a problem of observer selection effects and the "anthropic shadow" at play with global catastrophic risks — if nuclear wars have the potential to extinguish civilizations in possible worlds where they occur, then civilizations that exist to observe their histories are unlikely to have a good sense of the true frequency of nuclear wars.[19] To the extent that many nuclear wars are likely not extinction risks *per se* — as discussed below — this is not a strong objection to probabilistic estimates.[20] Nonetheless, like the winners of a coin-flipping

---

[16] Rodriguez, "How Likely Is a Nuclear Exchange between the US and Russia?" (Rethink Priorities), accessed February 14, 2023, https://rethinkpriorities.org/publications/how-likely-is-a-nuclear-exchange-between-the-us-and-russia.
[17] For detailed calculations, see Nuclear Probabilities.
[18] "This order-of-magnitude estimate of 1 percent per year includes a range from a third of a percent to 3 percent per year, but the risk is likely to be upper bounded by 10 percent per year since we have survived sixty-six years of nuclear deterrence without any use of nuclear weapons in war, much less a major exchange. Similarly, 0.1 percent per year is likely to be a lower bound on the risk since that would imply that current policies could be continued for approximately one thousand years before there would be a significant probability of civilization being destroyed. Over that time period, a simple statistical argument would predict fifteen major crises since there has been one in the last sixty-six years, namely the Cuban missile crisis of 1962, and 1,000/66 = 15 after rounding. In light of the risks during that crisis that are detailed in appendix A, it is likely that at least one of fifteen such crises would result in a nuclear war." Hellman, "Probabilistic Risk Assessment" in *On Assessing the Risk of Nuclear War,* 96.
[19] Milan M. Ćirković, Anders Sandberg, and Nick Bostrom, "Anthropic Shadow: Observation Selection Effects and Human Extinction Risks," *Risk Analysis* 30, no. 10 (2010): 1495–1506, https://doi.org/10.1111/j.1539-6924.2010.01460.x.
[20] To expand on this point, observer selection effects are most likely for true extinction risks, i.e. events after which there are no observers left. Insofar as we do not think nuclear war is likely to lead *directly* to extinction, the anthropic shadow does not apply. It may, however, suggest that worlds in which there are significantly larger arsenals, or in which there are non-nuclear strategic deterrents like strategic bioweapons, or in which initial use of nuclear weapons leads to a rapidly deteriorating security environment and in turn the emergence of doomsday cults or the use of extinction-level weapons, are less likely to be observed.

contest, we ought not assume that our good luck necessarily tells us much about the coin or about our coin-flipping skills.[21] Moreover, the estimates about the probability of the outbreak of nuclear war obscure important distinctions about *different kinds* of nuclear war and about escalation probability curves.[22] More detailed analyses simply do not exist and we ought to be suspicious of highly complex and formalized models of risk insofar as they are not grounded on reference class forecasting or other empirical bases. It may be best to follow a dictum of net assessment: "model simple and think complex."[23]

Uncertainty on Consequences

Second, what are the effects of nuclear war? Again, we do not really know.[24] In the aftermath of the atomic bombings of Japan, medical personnel collected information documenting the horrors of the nuclear weapons dropped on Hiroshima and Nagasaki: blast damage, flash burns, blindness, fire, acute radiation poisoning, and long-term health effects.[25] Nuclear testing later revealed other issues. Atomic scientists soon realized that nuclear fallout — radioactive material that is catapulted into the air and spread over the earth's surface during some nuclear explosions — could present a serious problem, as it did when radioactive ash rained on the tuna fishing boat *Lucky Dragon* after a 1954 U.S. nuclear test and poisoned the crew.[26] Similarly, the 1962 Starfish Prime test of a high-altitude nuclear detonation created an unexpectedly large electro-magnetic pulse (EMP) effect, revealing yet another danger from nuclear weapons for a modern civilization that relies on EMP-vulnerable critical infrastructures.[27]

---

[21] For a related discussion of the stock market as a coin-flipping contest, see Warren Buffet, "The Superinvestors of Graham-and-Doddsville," *Hermes, the Columbia Business School Magazine* (1984).
[22] See below, "Nuclear Wars Are Not Created Equal."
[23] Thanks to Matthew Gentzel for his help in understanding the applicability of net assessment to these problems. Paul Bracken, "Net Assessment: A Practical Guide," *The US Army War College Quarterly: Parameters* 36, no. 1 (March 1, 2006), https://doi.org/10.55540/0031-1723.2285.
[24] This section and the related points in the following sections are indebted to conversations with Dr. James Scouras and to Dr. Scouras's "Nuclear War as a Global Catastrophic Risk," *Johns Hopkins Applied Physics Laboratory,* which provides an excellent overview of the topic.
[25] For an accessible discussion, see The MIT Press Reader, "The Devastating Effects of Nuclear Weapons," *The MIT Press Reader* (blog), March 2, 2022, https://thereader.mitpress.mit.edu/devastating-effects-of-nuclear-weapons-war/.
[26] David Ropeik, "How the Unlucky Lucky Dragon Birthed an Era of Nuclear Fear," *Bulletin of the Atomic Scientists*, 2018, https://thebulletin.org/2018/02/how-the-unlucky-lucky-dragon-birthed-an-era-of-nuclear-fear/.
[27] "[S]ome phenomena were discovered late, and by surprise, in the nuclear test program. For example, an unexpectedly large EMP was observed in the Starfish Prime atmospheric nuclear test in 1962." James Scouras, ed., *On Assessing the Risk of Nuclear War* (Johns Hopkins Applied Physics Laboratory, 2021), 9. See also John S Foster et al., "Report of the Commission to Assess the Threat to the United States from Electromagnetic Pulse (EMP) Attack" (EMP Commission, 2004).

Nuclear testing did produce data on the observable physical effects of nuclear weapons, creating well-characterized understandings of blast radii and yields.[28] As the scholar of nuclear risk Dr. James Scouras points out, however, our knowledge about nuclear weapons is biased towards these easily-observable and -measurable physical effects.[29] One of the most significant effects of nuclear war is "nuclear winter": the climate effects of soot in the Earth's atmosphere following firestorms after a nuclear war. This remains poorly understood. As discussed below in The Non-Linearity of War Effects, several aspects of nuclear winter research — including uncertainties about black carbon release and transportation into the stratosphere — lead to extreme uncertainty surrounding possible climate effects.

Beyond nuclear winter, the societal consequences of nuclear war are even more uncertain. How would widespread grief and post-traumatic stress disorder affect recovery efforts? How would agricultural practices change in response to nuclear winter? Would there be food hoarding or global cooperation to share limited resources and avert mass starvation? Could liberal democracies withstand the challenges of major nuclear war better or worse than autocracies? Would nuclear use erode norms around other weapons of mass destruction, such as strategic biological weapons? Would it hasten the development and unsafe deployment of risky emerging technologies? At what point would global civilization collapse? How likely is recovery after a civilizational collapse? How likely is a recovery with *good values* after a civilizational collapse? Would a totalitarian hegemon emerge in the aftermath of nuclear war? Is total human extinction likely?

Again, it is possible to build complex models to try to answer these questions, and some researchers have made admirable efforts to understand civilizational collapse, but fundamentally we are speculating on the answers to these questions.[30]

Uncertainty on Risk Reduction

[This section is adapted from a previous Founders Pledge report on "Philanthropy to the Right of Boom." Readers familiar with that document may wish to skip to the section on Reasoning under Uncertainty.]

---

[28] These data inform tools like NukeMap, where users can visualize the theoretical direct effects of various nuclear weapons on the world map.

[29] "Somewhat less attention was paid to those phenomena that were inherently hard to predict or whose effects were delayed." James Scouras, ed., *On Assessing the Risk of Nuclear War* (Johns Hopkins Applied Physics Laboratory, 2021), 9.

[30] Luisa Rodriguez, "What Is the Likelihood That Civilizational Collapse Would Directly Lead to Human Extinction (within Decades)?," *Effective Altruism Forum*, 2020, https://forum.effectivealtruism.org/posts/GsjmufaebreiaivF7/what-is-the-likelihood-that-civilizational-collapse-would.

Third, what can we do about this risk? Once again, we do not really know. If we care about preventing all-out war between the United States and Russia, for example, what should we do? Should we fund track II (i.e. non-governmental) diplomatic dialogues to discuss the future of arms control after the demise of the New START agreement? Should we focus on understanding the effects of applications of new technologies such as artificial intelligence on strategic stability? Should we promote civil defense and agricultural resilience to prepare for worst case scenarios? Should we fund grassroots campaigns for nuclear arms control and disarmament?

Philanthropists may have some considerations that bear on these questions. For example, funders may believe that nuclear disarmament is an intractable goal, given the political and military realities of the world; or that the world currently represents one of the more stable distributions of capabilities. More fundamentally, however, we continue to have a poor understanding of the sources of nuclear risk, its probability, and its consequences.[31] Unlike other issue areas where the mechanism of change is clearer, scholars of nuclear war disagree on fundamental issues, such as whether a "no first-use" or "sole purpose" declaratory policy would be desirable.[32]

This uncertainty is not just the uncertainty of any non-expert funder, that could be resolved by learning more about the field. Subject-matter experts' theories remain untested and often untestable.[33] Historical accounts of states' behavior can provide some evidence, but these studies face the problem of the counterfactual. Probabilistic forecasting can help aggregate the "wisdom of the crowd" but, in questions about low-probability, high-consequence risks, these methods lack the objective scoring metrics that make them so powerful in other contexts; the data to resolve forecasts either do not exist or are too few.[34] Similarly, new attempts in international relations to construct "experimental

[31] This uncertainty is often not just about effect size, but also about effect *sign*. In the words of one analysis, "Benefit-cost analysis and other structured analytic methods applied to evaluate risk mitigation measures must acknowledge that we often do not even know whether many proposed approaches (e.g., reducing nuclear arsenals) will have a net positive or negative effect." (James Scouras, ed., *On Assessing the Risk of Nuclear War* (Johns Hopkins Applied Physics Laboratory, 2021), vii) As discussed below, some approaches like experimental wargaming may help with these methodological challenges.
[32] Galen Jackson et al., "Nuclear First-Use and Presidential Authority," *Texas National Security Review*, 2019, https://tnsr.org/roundtable/policy-roundtable-nuclear-first-use-and-presidential-authority/.
[33] The idea of an "inside view" model was first introduced by Kahneman in 1993: "An inside view forecast draws on knowledge of the specifics of the case, the details of the plan that exists, some ideas about likely obstacles and how they might be overcome. In an extreme form, the inside view involves an attempt to sketch a representative scenario that captures the essential elements of the history of the future. In contrast, the outside view is essentially statistical and comparative, and involves no attempt to divine future history at any level of detail." Daniel Kahneman and Dan Lovallo, "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking," *Management Science 39*, No. 1 (Jan., 1993): 25.
[34] For a discussion of this problem, see Ezra Karger, Pavel D. Atanasov, and Philip Tetlock, "Improving Judgments of Existential Risk: Better Forecasts, Questions, Explanations, Policies," *Future of Humanity*

wargames" appear promising, but also run into the problem of "ecological validity" — how well do the games actually represent the reality they purport to simulate?[35]

The problem, once again, is that the world has very limited experience with nuclear war and it is unclear how this experience translates to present challenges. Thus there is little opportunity for falsification of theories and there are limited reference classes and base rates for understanding nuclear war, the likelihood of escalation, and the consequences of different kinds of nuclear war. This allows for a wide range of plausible viewpoints and possible interventions for philanthropists.

These deep uncertainties also mean that we do not know much about what "safe" interventions might be. Later sections of this report suggest that interventions "right of boom" — after first nuclear use — may be high-impact funding opportunities: for example, developing a deeper understanding of escalation management. A common objection to this is that such interventions are dangerous because they may make war appear "winnable" and may thus raise the probability of nuclear war even if they mitigate the consequences of such a war. This is a fair concern, and one that ought to be studied further (by funding more policy-relevant right-of-boom research).

It is important to note, however, that philanthropists face similarly high uncertainty about the potential downsides of more ideologically palatable interventions, such as disarmament. At first glance, disarmament appears obviously good. The current distribution of capabilities, however, may represent one of the more stable configurations and possible worlds.[36] There has been no nuclear war since 1945, slow nuclear weapons proliferation, minimum deterrence arsenals in many nuclear states, restraint on many kinds of strategic weapons systems, and no all-out race on arsenal size for several decades.[37]

Whether nuclear deterrence has contributed to the "Long Peace" of the 20th and 21st centuries — and whether the Long Peace is a statistically meaningful phenomenon — is an

---

Institute Technical Report, 2022, https://doi.org/10.2139/ssrn.4001628. Note that this is not a claim about "black swans" or the supposed impossibility of predicting certain events. For a response to this separate issue, see Philip Tetlock, Yunzi Lu, and Barb Mellers, "False Dichotomy Alert: Improving Subjective-Probability Estimates vs. Raising Awareness of Systemic Risk," SSRN Scholarly Paper (Rochester, NY, January 20, 2022), https://doi.org/10.2139/ssrn.4013831.

[35] Andrew W Reddie and Bethany L Goldblum, "Evidence of the Unthinkable: Experimental Wargaming at the Nuclear Threshold," *Journal of Peace Research*, October 14, 2022, https://doi.org/10.1177/00223433221094734.; Erik Lin-Greenberg, Reid B.C. Pauly, and Jacquelyn G. Schneider, "Wargaming for International Relations Research," *European Journal of International Relations* 28, no. 1 (March 1, 2022): 83–109, https://doi.org/10.1177/13540661211064090.

[36] Thanks to Dr. Andrew Reddie for this point.

[37] Thanks to Matthew Gentzel for his insights on this.

issue under debate, but is a relevant factor for weighing the benefits of disarmament.[38] Smaller arsenals may, moreover, change targeting practices such that the expected cost increases — some arguments for "minimum deterrents" rely on the targeting of cities, civilian populations, and industrial centers, rather than counterforce targeting.[39] The key point is that uncertainty abounds on *all* risk reduction measures, including about the *sign* of interventions (i.e. whether it is net good or bad). This is discussed in greater detail below ("Is Disarmament Obviously Net-Positive?")

Reasoning Under Uncertainty: Impact Multipliers and Crucial Considerations

High uncertainty on probabilities, consequences, and risk reduction is not the same as cluelessness. As explained in Founders Pledge's *Guide to the Changing Landscape of High-Impact Climate Philanthropy,* "Even when we deal with large uncertainties, this does not mean we cannot make statements about the relative promise of different strategies and the likelihood that they will be highly impactful."[40]

As explained in the next two sections, there are some observable features of the structure of nuclear risk and the allocation of philanthropic funds that allow philanthropists to identify crucial considerations for strategic giving and develop "impact multipliers."

[The rest of this section is adapted from Founders Pledge's report on "Philanthropy to the Right of Boom." Readers familiar with this document may wish to skip to the next section, The Changing Landscape of Nuclear Risk Reduction.]

In this context, the term "impact multipliers" refers to features of the world that make one funding opportunity relatively more effective than another.[41] Stacking these multipliers makes effectiveness a "conjunction of multipliers;" understanding this conjunction can in

---

[38] For a discussion of the statistical meaning of the Long Peace, and the likelihood of observing such a period, see Bear F. Braumoeller, *Only the Dead: The Persistence of War in the Modern Age* (Oxford University Press, 2019).

[39] "Instead of simply threatening massive damage on a foe, proportional deterrence seeks to specifically communicate to adversaries that such destruction will quantitatively and qualitatively cancel any possible gains. This modern iteration of minimum deterrence shares a basic conceptual continuity with Brodie's depiction in that it hinges on the willingness of the deterring state to hold the "aggressor's population/industrial centers" at risk. Advocates of minimum deterrence describe this view as a "countervalue" approach that involves punishing one's adversary versus a "counterforce" approach that focuses on targeting an adversary's military in order to deny its wartime objectives." Joshua D Wiitala, "Challenging Minimum Deterrence: Articulating the Contemporary Relevance of Nuclear Weapons," *Air and Space Power Journal*, Spring 2016, 18.

[40] Johannes Ackva, Luisa Sandkühler, and Violet Buxton-Walsh, "A Guide to the Changing Landscape of High-Impact Climate Philanthropy" (Founders Pledge, November 2021), 55, https://founderspledge.com/stories/changing-landscape.

[41] For a discussion of impact multipliers, see Johannes Ackva, Luisa Sandkühler, and Violet Buxton-Walsh, "A Guide to the Changing Landscape of High-Impact Climate Philanthropy" (Founders Pledge, November 2021), 14-16, https://founderspledge.com/stories/changing-landscape.

turn help guide philanthropists seeking to maximize their impact under high uncertainty.[42] Doing this may not allow us to understand *absolute* cost-effectiveness (in terms of lives saved per dollar), but does allow us to understand *relative* cost-effectiveness, and thereby rank and identify the most effective interventions.[43]

Not all impact multipliers are created equal, however. To systematically engage in effective giving, philanthropists must understand the largest impact multipliers — "critical multipliers". These are features that most dramatically cleave more effective interventions from less effective interventions. In global health and development, for example, one critical multiplier is simply to focus on the world's poorest people. Because of large inequalities in wealth and the decreasing marginal utility of money, helping people living in extreme poverty rather than richer people in the Global North is a critical multiplier that winnows the field of possible interventions more than many other possible multipliers.

Additional considerations — the prevalence of mosquito-borne illnesses, the low cost and scalability of bednet distribution, and more — ultimately point philanthropists in global health and development to one of the most effective interventions to reduce suffering in the near term: funding the distribution of insecticide-treated bednets.[44] This report represents an attempt to identify defensible critical multipliers in nuclear philanthropy,[45] and potentially to move one step closer to finding "the nuclear equivalent of mosquito nets." [46]

There are many potential impact multipliers in nuclear philanthropy. For example, as explained below, focusing on states with larger nuclear arsenals is likely to be more impactful than focusing on nuclear terrorism. Nuclear terrorism would be horrific and a single attack in a city (e.g. with a dirty bomb) could kill thousands of people, injure many more, and cause long-lasting damage to the physical and mental health of millions.[47] All-out nuclear war between the United States and Russia, however, would be many times worse.

---

[42] Thomas Kwa, "Effectiveness Is a Conjunction of Multipliers," *Effective Altruism Forum*, 2022, https://forum.effectivealtruism.org/posts/GzmJ2uiTx4gYhpcQK/effectiveness-is-a-conjunction-of-multipliers.

[43] Appendix 2 nonetheless models a simplistic absolute cost-effectiveness and break-even analysis to illustrate the plausible cost-effectiveness of large investments in nuclear risk reduction.

[44] GiveWell, "Against Malaria Foundation," accessed February 14, 2023, https://www.givewell.org/charities/amf.

[45] And possibly great power conflict and international security more broadly.

[46] This phrase is from Jeffrey Lewis's appearance on the 80,000 Hours podcast, Rob Wiblin, "Jeffrey Lewis on the Most Common Misconceptions about Nuclear Weapons," accessed February 14, 2023, https://80000hours.org/podcast/episodes/jeffrey-lewis-common-misconceptions-about-nuclear-weapons/#robs-intro-000000.

[47] The societal consequences of e.g. a dirty bomb attack may be more larger than the direct deaths, although the point about relative scale of the problems stands. Thanks to Dr. Andrew Reddie for this point in a round of reviews.

Hundreds of millions of people could die from the direct effects of a war. If we believe nuclear winter modeling, moreover, there may be many more deaths from climate effects and famine than from the blast and other direct impacts of the bombs.[48] In the worst case, civilization could collapse. Simplifying these effects, suppose for the sake of argument that a nuclear terrorist attack could kill 100,000 people, and an all-out nuclear war could kill 1 billion people. *All else equal*, in this scenario it would be 10,000 times more effective to focus on preventing all-out war than to focus on nuclear terrorism.[49]

Generalizing this pattern, philanthropists ought to prioritize the largest nuclear wars (again, all else equal) when thinking about additional resources *at the margin*. This can be operationalized with real numbers — nuclear arsenal size, military spending, and other measures can serve as proxy variables for the severity of nuclear war, yielding rough multipliers.[50] These measures are imperfect proxies — nuclear weapons can be viewed as a way to "offset" costly conventional military spending, as in the Eisenhower administration's First Offset strategy, which emphasized more "bang for the buck" — and should not be used in isolation.[51] This could lead philanthropists to prioritize risk mitigation around adversarial relationships between the world's "great powers." Similarly, because risk is a function of probability and consequence, philanthropists who aspire to be most effective can prioritize the nuclear relationships that are most likely to lead to war (and, further, those that are most likely to draw in larger powers).

These impact multipliers already yield useful conclusions. They suggest a focus on the behavior of major nuclear-armed states and, within this, a focus on preventing the largest

---

[48] Models of the climatic effects of nuclear war — especially those that predict severe and prolonged global cooling in a "nuclear winter" — are controversial and highly politicized. For a glimpse of the debate, see the exchange between Reisner *et al.* and Robock *et al.* in *Atmospheric Chemistry and Physics* and *JGR Atmospheres* (A. Robock et al., "Climatic Consequences of Regional Nuclear Conflicts," *Atmospheric Chemistry and Physics* 7, no. 8 (April 19, 2007): 2003–12, https://doi.org/10.5194/acp-7-2003-2007.; Jon Reisner et al., "Climate Impact of a Regional Nuclear Weapons Exchange: An Improved Assessment Based On Detailed Source Calculations," *Journal of Geophysical Research: Atmospheres* 123, no. 5 (2018): 2752–72, https://doi.org/10.1002/2017JD027331; Alan Robock, Owen B. Toon, and Charles G. Bardeen, "Comment on 'Climate Impact of a Regional Nuclear Weapon Exchange: An Improved Assessment Based on Detailed Source Calculations' by Reisner et Al.," *Journal of Geophysical Research: Atmospheres* 124, no. 23 (2019): 12953–58, https://doi.org/10.1029/2019JD030777; Jon Reisner et al., "Reply to Comment by Robock et al. on 'Climate Impact of a Regional Nuclear Weapon Exchange: An Improved Assessment Based on Detailed Source Calculations,'" *Journal of Geophysical Research: Atmospheres* 124, no. 23 (2019): 12959–62, https://doi.org/10.1029/2019JD031281)
[49] All else, obviously, is not equal. Questions about the tractability of escalation management are crucial. This is discussed in greater detail below.
[50] This is more difficult than just counting warheads. See David C. Logan, "The Nuclear Balance Is What States Make of It," *International Security* 46, no. 4 (April 1, 2022): 172–215, https://doi.org/10.1162/isec_a_00434.
[51] Thanks to Matthew Gentzel for this point.

wars. This may differ substantially from the approach of some traditional philanthropic actors working on nuclear security. We expand on these points in greater detail below.

The next section discusses the funding patterns of these actors in light of the changing threat landscape of the so-called "New Nuclear Age." This discussion informs the search for impact multipliers in the sections that follow later in the report.

Questions for Further Investigation

- What are the closest reference classes for forecasting the probability of nuclear war?
- How can scholars establish the ecological validity (or invalidity) of synthetic data generation methods, including experimental wargaming and second-generation probabilistic forecasting?
- What can natural catastrophes with solar radiation-blocking mechanisms that are similar to "nuclear winter" — including large-magnitude explosive volcanic eruptions and asteroid collisions — teach us about the possible global effects of nuclear war?
- Can better knowledge of the *consequences* of nuclear war improve our understanding of the *probabilities* of nuclear war? (e.g., by removing doubts about observer selection effects)

The Changing Landscape of Nuclear Risk Reduction

Key Points:

- China's growing nuclear ambitions, coupled with the influence of emerging technologies, may create unstable new dynamics.
- The MacArthur Foundation's withdrawal from nuclear security and the collapse of philanthropic entities associated with FTX have led to a severe funding shortfall.
- These dynamics may have created a moment of high philanthropic leverage over the field of nuclear security.

This section outlines key features of the landscape of nuclear risk and nuclear philanthropy. In short, a large philanthropic funding gap has opened at an inconvenient time. Structural features of the world — including heightened great power competition, a multipolar nuclear order, the breakdown of traditional arms control, and rapid advances in emerging technologies — may increase nuclear risk, or at least increase uncertainty about the magnitude of that risk.

At the same time, recent large-scale funding shortfalls in nuclear philanthropy and in catastrophic risk reduction broadly have left the field scrambling for money. On the one hand, this is highly unfortunate, especially in light of the potentially increased risk during the Russo-Ukrainian war (ongoing as of this writing in 2023). On the other hand, it presents a potentially high-leverage opportunity, as grantees may be more receptive to new ideas and approaches.

The "New Nuclear Age"?

Subject-matter experts in nuclear security so often refer to a "new nuclear age" that the term has become cliché.[52] Usually, this phrase refers to an interconnected set of

[52] Andrew F. Krepinevich, Jr., "The New Nuclear Age," *Foreign Affairs*, 2022, https://www.foreignaffairs.com/articles/china/2022-04-19/new-nuclear-age; Robert Legvold and Christopher F. Chyba, "Meeting the Challenges of a New Nuclear Age," *Daedalus* (2020) accessed February 2, 2023, https://www.amacad.org/daedalus/meeting-challenges-new-nuclear-age; The Economist, "'The World Is Entering a New Nuclear Age'—an Old Fear Returns," *The Economist*, https://www.economist.com/podcasts/2023/01/02/the-world-is-entering-a-new-nuclear-age-an-old-fear-returns; Rebecca Hersman, "Wormhole Escalation in the New Nuclear Age," Texas National Security Review, July 9, 2020, https://tnsr.org/2020/07/wormhole-escalation-in-the-new-nuclear-age/; *The Fragile Balance of Terror: Deterrence in the New Nuclear Age* (Cornell University Press, 2022), https://www.jstor.org/stable/10.7591/j.ctv310vm0j; "Deterrence, Norms, and the Uncomfortable Realities of a New Nuclear Age," War on the Rocks, April 20, 2020,

observations about international security in the 2020s. The first is the idea of a "multipolar nuclear order." As described in the recent book *The Fragile Balance of Terror*,

> "[T]he emergence of multiple nuclear states [since the Cold War] makes balances of power more complex and deterrence relationships more uncertain. Our theories and understanding derived from the Cold War bipolar nuclear competition leave us ill-equipped to handle the daunting challenges of this new nuclear age."[53]

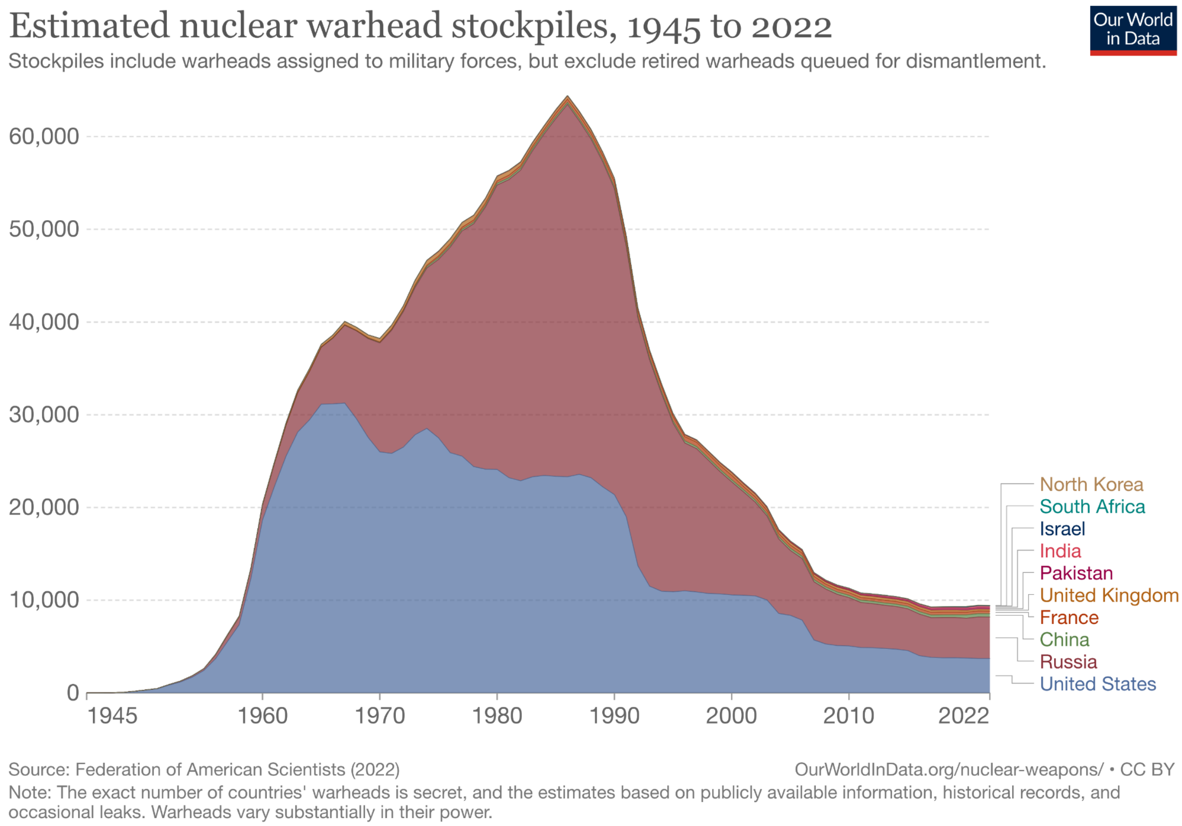
The question of whether more nuclear states make the world more or less safe does not, however, have an obvious answer, and there are arguments that multipolar deterrence is more stable.[54] Once again, philanthropists ought to be epistemically humble about what they can and cannot know about nuclear risk. We should remember, moreover, that most nuclear weapons states represent a tiny fraction of the world's arsenals. The United States and Russia continue to dominate these measures:

---

https://warontherocks.com/2020/04/deterrence-norms-and-the-uncomfortable-realities-of-a-new-nuclear-age/.

[53] Vipin Narang and Scott D. Sagan, "Introduction" in *The Fragile Balance of Terror: Deterrence in the New Nuclear Age.*

[54] Kenneth N. Waltz, "The Spread of Nuclear Weapons: More May Be Better: Introduction," *The Adelphi Papers* 21, no. 171 (September 1981): 1–1, https://doi.org/10.1080/05679328108457394.



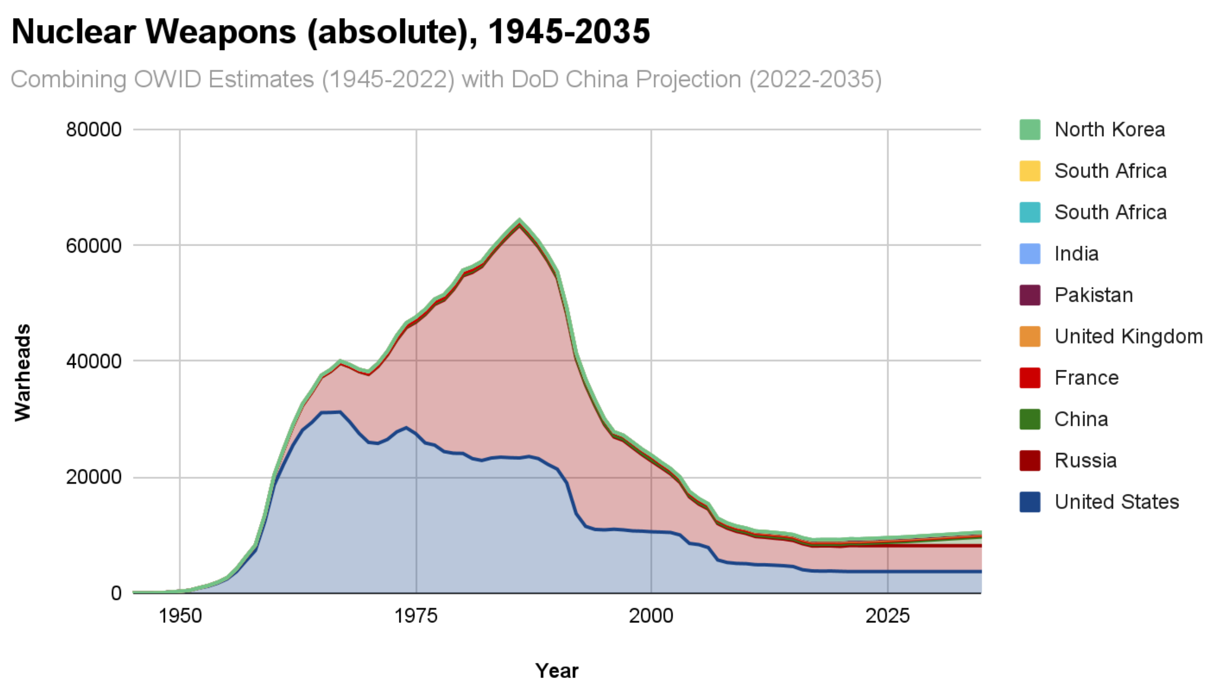
Source: Our World in Data, using data from *Federation of American Scientists*.

These reductions are even more significant than they look by warhead count; weapon yields were also decreasing dramatically with the removal of multi-megaton warheads from arsenals.[55] In general, assessing both absolute and relative nuclear capabilities is far more complicated than mere warhead counts can illustrate, with additional considerations including not only yield but also reliability.[56]

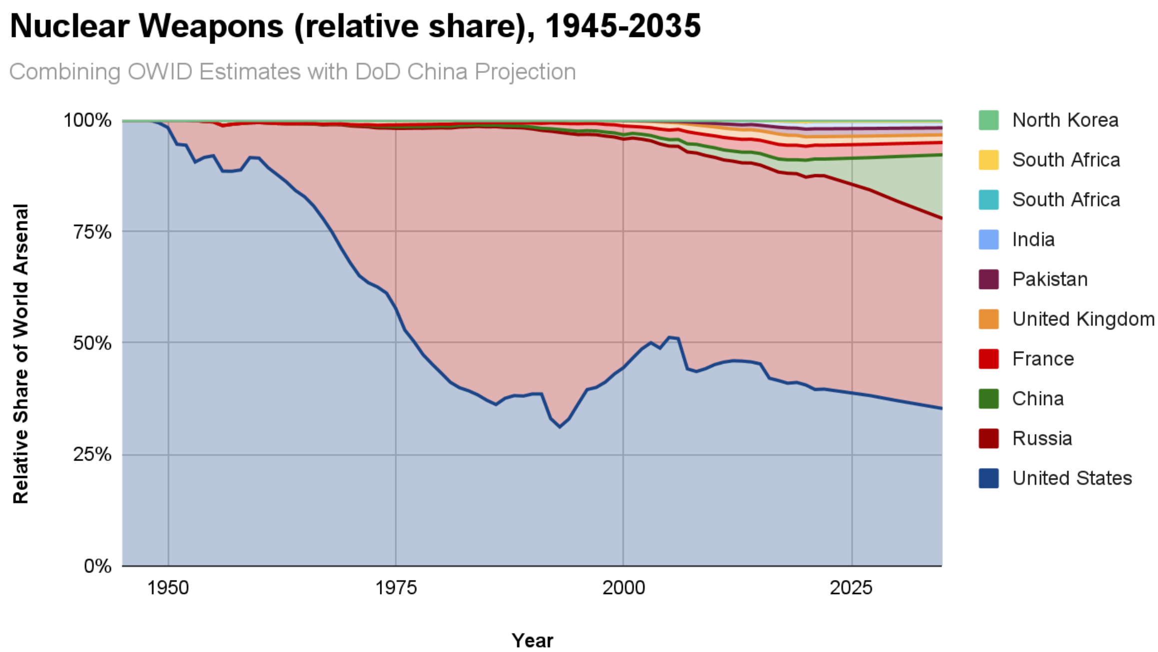
One emerging exception to the dominance of the United States and Russia is China, the third-largest nuclear power and one of the United States's main strategic rivals. In late 2022, the U.S. Department of Defense suggested in its *China Military Power Report* that "If China continues the pace of its nuclear expansion, it will likely field a stockpile of about 1500 warheads by its 2035 timeline."[57] If we take this estimate at face value, and for

[55] Thanks to Matthew Gentzel for this point.
[56] David C. Logan, "The Nuclear Balance Is What States Make of It," International Security 46, no. 4 (April 1, 2022): 172–215, https://doi.org/10.1162/isec_a_00434.
[57] U.S. Department of Defense, "2022 Report on Military and Security Developments Involving the People's Republic of China," November 2022, 94, https://www.defense.gov/News/Releases/Release/Article/3230516/2022-report-on-military-and-security-developments-involving-the-peoples-republi/https%3A%2F%2Fwww.defense.gov%2FNews%2FReleases%2FR

simplicity assume that other stockpiles will remain unchanged, China's arsenal is still small compared to U.S.-U.S.S.R. Cold War heights, but becomes an appreciable fraction of the world total by 2035:



elease%2FArticle%2F3230516%2F2022-report-on-military-and-security-developments-involving-the-peoples-republi%2F.



Source: Author's visualizations based on Our World in Data illustration (above), combining Our World in Data dataset and the U.S. Department of Defense China Military Power Report projection (2022).

Warhead number is not the only salient feature of China's changing nuclear ambitions.[58] Additional developments include building out a nuclear triad, testing a "fractional orbital bombardment system" with a hypersonic glide vehicle in 2021 (though China denies this was the purpose of the test in question), building 300 new missile silos, and potentially even changing fundamental parts of Chinese nuclear posture and policy.[59] In short, China's nuclear arsenal and policy are changing drastically and quickly.

The Three Body Problem

The features discussed above lead to another fact of the "new nuclear age" — renewed great power competition and the problem of "three-way deterrence" between Russia, China, and the United States. The broader problem of conflict between major military powers is discussed in Founders Pledge's report on *[Great Power Conflict](#)*. U.S. national security strategy documents also emphasize great power competition, but Founders

[58] David C. Logan, "The Nuclear Balance Is What States Make of It," *International Security* 46, no. 4 (April 1, 2022): 172–215, https://doi.org/10.1162/isec_a_00434.

[59] See Figure 1, "Summary of important changes to China's nuclear arsenal and policies during the last five years" in Jacob Stokes, "Atomic Strait: How China's Nuclear Buildup Shapes Security Dynamics with Taiwan and the United States" (Center for a New American Security, 2023), 3.

Pledge's focus on the issue stems from a desire to prevent human suffering and catastrophe *globally*, not from a pursuit of national interests. The 2022 National Defense Strategy lists great power competition as part of several "top-level priorities": "Defending the homeland, paced to the growing multi-domain threat posed by the People's Republic of China (PRC)" and "Deterring aggression, while being prepared to prevail in conflict when necessary – prioritizing the PRC challenge in the Indo-Pacific region, then the Russia challenge in Europe."[60] Similarly, the 2022 Nuclear Posture Review singles out the PRC as a primary factor in U.S. evaluation of nuclear deterrence:

> "The People's Republic of China (PRC) is the overall pacing challenge for U.S. defense planning and a growing factor in evaluating our nuclear deterrent. The PRC has embarked on an ambitious expansion, modernization, and diversification of its nuclear forces and established a nascent nuclear triad. The PRC likely intends to possess at least 1,000 deliverable warheads by the end of the decade."[61]

Competition with a near-peer competitor (the USSR) drove U.S. nuclear policy during the Cold War, too, but the challenge of "three-way deterrence" with two near-peer competitors is new.[62] Cold War game-theoretic models were designed largely with one major adversary in mind, but the 2022 Nuclear Posture Review makes it clear that the U.S. now believes it will need to deter two major nuclear powers.[63] Some of these issues are structural: negotiations become more complex when more than two parties must agree. Other issues may relate to targeting policy: new and challenging considerations arise over counter-force targeting and what kinds of deterrence regimes appear most desirable to states in three-way deterrence.[64] Not all changes are necessarily destabilizing:[65]

---

[60] "Department of Defense Releases Its 2022 Strategic Reviews" U.S. Department of Defense, accessed February 6, 2023, https://www.defense.gov/News/Releases/Release/Article/3201683/department-of-defense-releases-its-2022-strategic-reviews-national-defense-stra/https%3A%2F%2Fwww.defense.gov%2FNews%2FReleases%2FRelease%2FArticle%2F3201683%2Fdepartment-of-defense-releases-its-2022-strategic-reviews-national-defense-stra%2F.

[61] U.S. Department of Defense, *2022 Nuclear Posture Review,* 4, https://media.defense.gov/2022/Oct/27/2003103845/-1/-1/1/2022-NATIONAL-DEFENSE-STRATEGY-NPR-MDR.PDF.

[62] The Economist, "How Will America Deal with Three-Way Nuclear Deterrence?," *The Economist*, accessed February 14, 2023, https://www.economist.com/united-states/2022/11/29/how-will-america-deal-with-three-way-nuclear-deterrence.

[63] "By the 2030s the United States will, for the first time in its history, face two major nuclear powers as strategic competitors and potential adversaries. This will create new stresses on stability and new challenges for deterrence, assurance, arms control, and risk reduction." U.S. Department of Defense, *2022 Nuclear Posture Review*, 4.

[64] Thanks to Matthew Gentzel for this point.

[65] Thanks to Matthew Gentzel for these points.

- "Truels" (i.e., single-shot duels between three players) may further disincentivize first strikes, depending on the balance of capabilities;
- The existence of an additional large nuclear power may make a single coercive Russian nuclear monopoly less likely in certain catastrophic scenarios;
- These dynamics may make it easier in some situations for a third party to de-escalate crises between the other two.

Again, we do not know how these considerations will stack up; uncertainty is high. For these reasons, leaders of U.S. Strategic Command have expressed concern about what Admiral Charles Richards has compared to the "three-body problem" in physics, stating that "I'm not sure what strategic stability looks like in a three-party world," that STRATCOM has been "furiously" working on renewed deterrence theory, and that he remains deeply pessimistic despite this work: "There are exactly zero stable [...] three-body orbital regimes." [66]

Emerging Technologies

A third feature of the "New Nuclear Age" is the impact of new and emerging technologies on the nuclear balance. Not all of these developments are equally concerning, and the threat of some technologies has been exaggerated for political purposes. Discussion of the so-called "hypersonic arms race," for example, seems to misunderstand basic facts about the state of missile defense technology — some of it, like homeland missile defense, is in its infancy — and about the purpose of U.S. homeland missile defenses (they are designed with North Korea and Iran, not China and Russia, in mind).[67] (These considerations may be

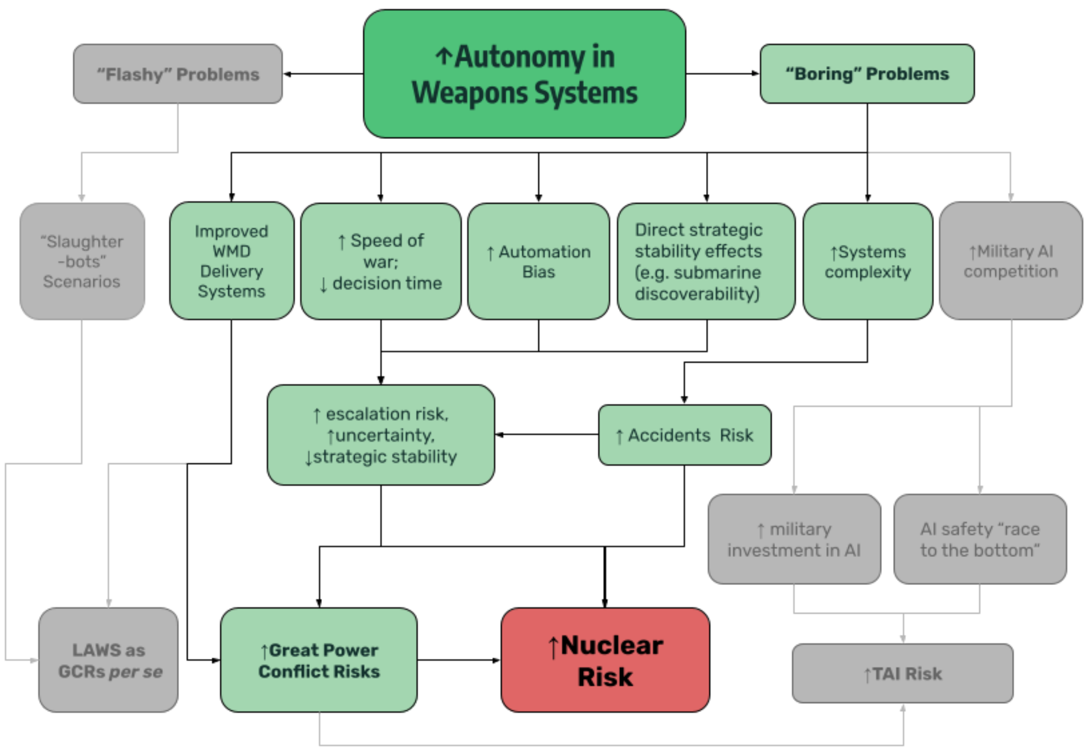
[66] "I'm not sure what strategic stability looks like in a three-party world, and a lot of terms have been kicked around. Oh, that's stabilizing, that's not stabilizing. That's destabilizing;" "I do know that there are many passively stable two body orbital regimes that you can stick stuff in, but there are exactly zero passively stable three body orbital regimes. They all require active stabilization. And I don't even know what that means when the forces can't be described by physics but are political, so we have gotten to think through this much harder than we have in the past;" "The free world is getting tested in ways right now, ways we haven't seen in decades. And that three-party nuclear-fueled world's just unprecedented." (STRATCOM, "2022 Space and Missile Defense Symposium," U.S. Strategic Command, August 11, 2022, https://www.stratcom.mil/Media/Speeches/Article/3126694/2022-space-and-missile-defense-symposium/https%3A%2F%2Fwww.stratcom.mil%2FMedia%2FSpeeches%2FArticle%2F3126694%2F2022-space-and-missile-defense-symposium%2F and Theresa Hitchens, "The Nuclear 3 Body Problem: STRATCOM 'furiously' Rewriting Deterrence Theory in Tripolar World," *Breaking Defense* (blog), August 11, 2022, https://breakingdefense.sites.breakingmedia.com/2022/08/the-nuclear-3-body-problem-stratcom-furiously-rewriting-deterrence-theory-in-tri-polar-world/).

[67] Andrew W. Reddie, "Hypersonic Missiles: Why the New 'Arms Race' Is Going Nowhere Fast," *Bulletin of the Atomic Scientists*, January 13, 2020, https://thebulletin.org/2020/01/hypersonic-missiles-new-arms-race-going-nowhere-fast/.

different for regional systems like THAAD and Aegis, where the effects of hypersonics may be more important.[68])

Other new technologies, however, are more concerning. Among these are applications of machine learning technologies to various systems, both nuclear and conventional, and the instability that this may create. Founders Pledge's 2022 report *Autonomous Weapons Systems and Military AI* discusses these risks in detail under the section titled "Pathways to Risk." The pathways to nuclear risk are outlined in the figure below, adapted from the same report:

Autonomy in Weapon Systems and Nuclear Risk



Source: Adapted from Founders Pledge report, *Autonomous Weapon Systems and Military AI* ("Pathways to Risk"). N.B. This merely purports to represent *possible, not* probable, pathways.

The list of potential new and emerging technologies that may affect strategic stability is long, and the signs of their effects (whether they will be positive or negative contributors to

[68] Thanks to Matthew Gentzel for emphasizing this distinction.

stability) are often unclear.[69] It is possible that applications of AI in nuclear command, control, and communications (NC3) — managed appropriately — can help improve threat assessment, decrease accident risk, and lengthen decision windows in ways that reduce risk overall. One potentially dangerous application of predictive machine learning systems is for "strategic warning" of nuclear attack.[70] The Department of Defense *Dictionary of Military and Associated Terms* no longer includes a definition of strategic warning, but used to define it as "a notification that enemy-initiated hostilities may be imminent" and defined "tactical warning" in contrast as "a notification that the enemy has initiated hostilities."[71] The incentives to adopt such systems are clear; if it provides early warning of an adversary's intended action, it provides early opportunity to prepare for that action. In the 1980s, the Soviet Union RYaN program (short for *Raketno-Yadernoe Napadenie,* or Nuclear Missile Attack) sought to quantify risk based on an aggregate of supposed strategic warning indicators dreamed up by Soviet leadership (such as "the amount of blood held in Western blood banks and the location of Western leaders").[72] Because there have been no examples of nuclear surprise attacks to train on, however, poorly implemented modern AI-enabled RYaN programs may be biased and increase the risk of war by giving false warnings.[73] Again, however, analysts ought to be epistemically humble about these issues; just because they *could* increase the risk of war does not mean that they will.

There are large uncertainties and new moving parts to this age, such that nuclear risk subjectively *feels* high. On February 21, 2023, Vladimir Putin announced that Russia would suspend its participation in the New START arms control agreement between Russia and

---

[69] "[T]he list of new technologies that can be identified as having potential significant consequences for strategic stability is long. These include broadly applicable enabling technologies such as artificial intelligence (AI), biotechnology (especially genetic engineering and synthetic biology), and quantum computing and cryptography. They include categories of counterspace weapons encompassing kinetic weapons, non-kinetic physical weapons (high-powered lasers and microwaves), cyber weapons, and electronic jamming and spoofing. They also include weapons whose characteristics might appear to an adversary as suited for executing first strikes, such as conventional and nuclear hypersonic weapons, including hypersonic glide vehicles (HGVs), hypersonic cruise missiles (HCMs), and stealthy strategic autonomous systems. And they include systems or capabilities that could help enable first strikes, such as persistent surveillance technologies for tracking mobile missiles, antineutrino detectors for tracking submerged SSBNs, and some aspects of counterspace and cyber weapons. There are also technologies that could in principle alter the underpinnings of multilateral strategic relationships, such as laser isotope separation for uranium enrichment." Christopher F Chyba, "New Technologies & Strategic Stability," *Daedalus*, accessed February 14, 2023, https://www.amacad.org/publication/new-technologies-strategic-stability.

[70] Thanks to Dr. James Scouras for pointing me to this threat and the example of RYaN.

[71] Cited in Ralph Strauch, *Strategic Warning and General War: A Look at the Conceptual Issues,* RAND N-1180-1-AF (1979), 5.

[72] Paul Scharre, *Four Battlegrounds: Power in the Age of Artificial Intelligence,* 287. For vivid descriptions of RYaN, see David E. Hoffman, *The Dead Hand: The Untold Story of the Cold War Arms Race and Its Dangerous Legacy.*

[73] For a deeper discussion of this, see Scharre, *Four Battlegrounds*, 287-289.

the United States.[74] The extended agreement was set to expire in 2026, already creating an uncertain future for strategic arms control.[75]

Whether the risk is — as the 2023 state of the "Doomsday Clock" at "90 seconds to midnight" implies — as high as it was at the height of the Cold War, when fewer risk reduction measures were in place and U.S. and Russian arsenals were much larger, is unclear.[76] At the same time, however, current philanthropic funding for nuclear risk reduction is unusually low, creating a danger of neglect just as the world is changing; the next section discusses this funding shortfall.

Funding Shortfalls: MacArthur and FTX

In 2022, the Founders Pledge Patient Philanthropy Fund made a grant in support of an analysis by Alex Toma of the Peace and Security Funders Group on *The Changing Landscape of Nuclear Security Philanthropy: Risks and Opportunities in the Current Moment*. The analysis can be found here.

To summarize the key points of the report:

- In 2021, the MacArthur Foundation announced that it would withdraw its support of nuclear security grantmaking via a $30 million capstone project distributed over multiple recipients — final funds will be disbursed in 2023.[77]
- MacArthur was the largest private funder of nuclear security work. Its withdrawal was described as "a big blow" to arms control.[78]
- This funding gap may be partially filled by the entry of new funders.
- Not all of this is bad, and new funders may have the opportunity to reshape the field in beneficial ways.

Since the publication of Toma's analysis, the behavior of another funder has changed this picture. In late 2022, various corporate and philanthropic entities associated with FTX and

[74] Heather Williams, "Russia Suspends New START and Increases Nuclear Risks," *Center for Strategic and International Studies*, February 23, 2023, https://www.csis.org/analysis/russia-suspends-new-start-and-increases-nuclear-risks.
[75] Department of State, "New START Treaty," United States Department of State, accessed February 14, 2023, https://www.state.gov/new-start/.
[76] Bulletin of the Atomic Scientists, "It is 90 Seconds to Midnight: 2023 Doomsday Clock Statement," *Bulletin of the Atomic Scientists*, accessed 28 February 2023, https://thebulletin.org/doomsday-clock/current-time/.
[77] MacArthur Foundation, "Nuclear Challenges Program Strategy," accessed February 14, 2023, https://www.macfound.org/programs/nuclear/strategy.
[78] Bryan Bender, "'A Big Blow': Washington's Arms Controllers Brace for Loss of Their Biggest Backer," *POLITICO*, July 19, 2021, https://www.politico.com/news/2021/07/19/washington-arms-controllers-nuclear-weapons-500126.

Sam Bankman-Fried collapsed and filed for bankruptcy. Grants committed by associated entities were reportedly not distributed and may be at risk of clawback in bankruptcy court.[79] Archived pages of the now-defunct FTX Foundation's Future Fund's website show that one of the single largest grants was made in part to support nuclear grantmaking, and that global catastrophic risks and great power conflict would have been major areas of focus for future grants.[80] The effect of the FTX collapse on this grantmaking remains uncertain, but it appears to have shrunk the pool of possible funding for nuclear security significantly.[81]

Without new funders, the state of philanthropic funding for nuclear security looks even more dire than it did in early 2022. On average, MacArthur made grants of around $15 million per year between 2014 and 2020. Total philanthropic nuclear security funding stood at about $47 million per year, such that the field faces a **32% reduction in funding.**[82] The centrality of MacArthur to the field is visualized in Toma's report — all the organizations in MacArthur's "orbit" are at risk of running out of funding:

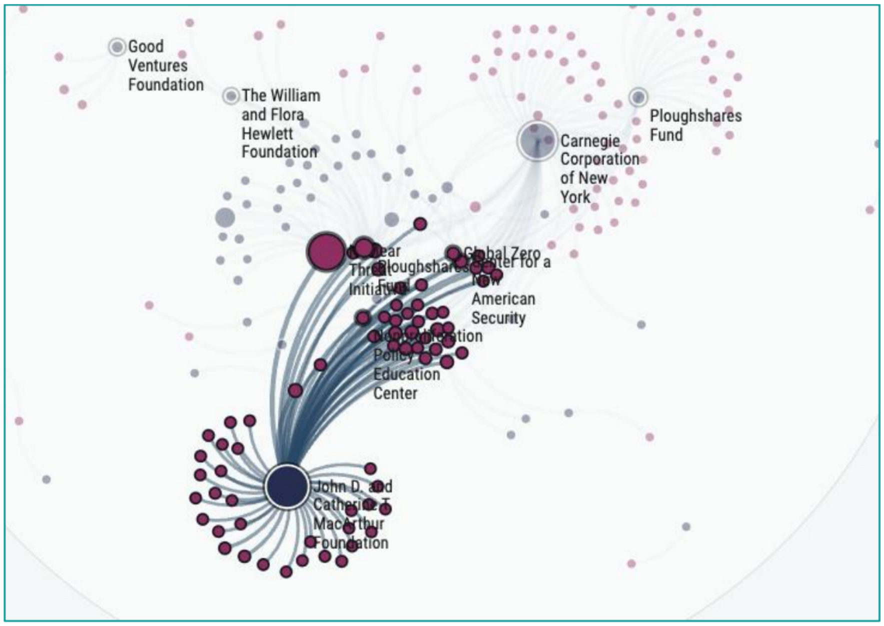
---

[79] I am not a lawyer and this is not legal advice.

[80] In February 2022, the Future Fund made a $15,000,000 grant recommendation to "support Longview's independent grantmaking on global priorities research, nuclear weapons policy, and other longtermist issues." ("Our Grants," FTX Future Fund [archived web page, 30 November 2022])

[81] This can be gleaned in part from the following edit, indicated in brackets: "The co-leads will make grants potentially totalling up to $10 million initially, a figure which could grow substantially if they find or create sufficiently strong opportunities. [Update December 2022: This is now to be determined as we seek new funders for this work.]" Kit Harris, "New Nuclear Security Grantmaking Programme at Longview Philanthropy," Effective Altruism Forum, accessed February 14, 2023, https://forum.effectivealtruism.org/posts/M7wNHbpqnLfDzmDK9/new-nuclear-security-grantmaking-programme-at-longview.

[82] Alexandra Toma, "The Changing Landscape of Nuclear Security Philanthropy: Risks and Opportunities in the Current Moment," https://static1.squarespace.com/static/62435b2d773bcf0ad9cbb672/t/62cee05cb38a4b2487a533e0/1657725020334/The+Changing+Landscape+of+Nuclear+Security+Philanthropy.pdf.



Source: Alex Toma, *The Changing Landscape of Nuclear Security Philanthropy,* from Peace and Security Funding Index visualization.

Bad Timing, High Leverage

In its coverage of the MacArthur withdrawal, *Vox* called this the “worst possible time” for a shortfall in nuclear security philanthropy, in light of the Russo-Ukrainian War.[83] While it is indeed bad timing, this moment may also represent a time of unusually high leverage over the field. As Toma pointed out, “amidst the turmoil of this major funder leaving, there is opportunity to rethink and reshape the field.”[84] This leverage is even higher in the wake of the FTX collapse. The field may be more open to a reassessment of strategic priorities than it otherwise would be. The next section outlines structural features of the problem of nuclear risk that can guide impact-oriented philanthropists in developing these priorities.

Questions for Further Investigation

[83] Dylan Matthews, “The Biggest Funder of Anti-Nuclear War Programs Is Taking Its Money Away,” Vox, March 17, 2022, https://www.vox.com/2022/3/17/22976981/nuclear-war-russia-ukraine-funding-macarthur-existential-risk-effective-altruism-carnegie.

[84] Alexandra Toma, “The Changing Landscape of Nuclear Security Philanthropy: Risks and Opportunities in the Current Moment,” https://static1.squarespace.com/static/62435b2d773bcf0ad9cbb672/t/62cee05cb38a4b2487a533e0/1657725020334/The+Changing+Landscape+of+Nuclear+Security+Philanthropy.pdf.

- Are there strong reasons to believe that the Department of Defense's assessment of Chinese nuclear ambitions is misguided or mistaken?
- What can policymakers and philanthropists learn from the stability of previous international orders with multiple great powers (e.g. Europe in the 19th Century)?
- What are other ways that nuclear risk and AI risk interact?
- What technologies *other than AI* pose major risks to nuclear stability? How might advances in space technology, synthetic biology, cyber operations, and non-nuclear strategic weapons generally affect nuclear risk?
- How likely are other major philanthropists to fill the gaps left by MacArthur and FTX?

The Structure of the Problem

Key Points:

- Large nuclear wars are disproportionately worse than limited nuclear use;
  - This holds from a near-term perspective, but becomes especially pronounced when factoring in existential risks such as unrecovered civilizational collapse.
- It is unclear whether limited violation of the "nuclear taboo" would weaken or strengthen the norm of non-use; the consequences of norm violation (punishment) matter.
- Funders face deep uncertainty about the effectiveness of different kinds of interventions (expanding on discussion above).
- The history of "close calls" suggests that there is a real risk of accidents and unintended nuclear escalation.
- Cold War arsenal reductions may be undone and the possibility of arms races should not be dismissed.

The previous section framed the current moment in nuclear risk and philanthropic support of risk-reduction measures. To summarize:

- There are large uncertainties about the magnitude of future nuclear risk, especially around Chinese intentions and nuclear ambitions.
- There has been a major funding shortfall, such that nuclear issues are unusually neglected.
- This may represent a moment of high philanthropic leverage over the field.

This section discusses key facts about the *problem structure* of nuclear risk. The aim is to reason from simple first principles, rather than try to build complex and untestable models about deterrence and arms control, and should be read in the context of philanthropic giving. Several key facts about the structure of the problem emerge:

1. **Nuclear wars are not created equal** — *all else equal*, if a given grant prevents a large nuclear war it is more cost effective than if it prevents a small nuclear war.
2. There is a **superlinear relationship between war size and global war costs (in human wellbeing)** — larger wars are disproportionately worse than smaller wars.

3. Funders, experts, and decision-makers face **deep uncertainty about the effectiveness of interventions** — we often simply do not know what would work best, and have no way of finding out.
4. **Accidents happen** — the problems of nuclear crisis management and escalation control are likely here to stay, and cannot be ignored.
5. **What comes down can go back up** — Cold War arsenal reductions are not guaranteed to be "sticky," and some trends suggest that states are interested in arming; the magnitude of nuclear risk could increase significantly in the near future.

Nuclear Wars Are Not Created Equal

Some nuclear wars — such as an all-out thermonuclear war between two states with arsenals the size of the Cold War United States and Russia — could be civilization-ending events, especially but not only if we give credence to high estimates of the probability and severity of the climate effects of nuclear winter (discussed below). On the other hand, the Second World War was technically a nuclear war, but that nuclear use was not a civilization-ending event, despite its horrific consequences for the inhabitants of Hiroshima and Nagasaki.[85]

There is, in short, a spectrum of possible nuclear wars. At the lowest end of this spectrum (with a questionable use of the word "war") is dirty-bomb terrorism, whose direct effects are highly limited, and which is designed to incite fear in a population, rather than to maximize damage and death. At the highest end of the spectrum exist possible futures (discussed below) where many states deploy extremely large arsenals and a global conflagration could kill billions of people. Indeed, some early Cold War strategists believed that the distinction between "limited" and "all-out" nuclear war is more salient than the distinction between "nuclear" and "non-nuclear" war.[86]

This fact matters because philanthropists have limited resources and must therefore prioritize both between and within issue areas. Funders therefore need to decide where to focus their efforts; "preventing the use of nuclear weapons" may be too broad a concept to

---

[85] Centeno *et al.* argue that the world wars did technically constitute a civilizational collapse, but clearly not on the scale discussed in this report: "The 'Second 30 Years War' (1914-1945) [...] did produce a collapse in the global social, and political order (Ferguson, 2006). This last example also serves as a reminder that one person's collapse may be another's opportunity; in the case of barbaric regimes, collapse is widely welcomed by the oppressed." (Centeno *et al.*, "Globalization and Fragility: A Systems Approach to Collapse" in *How Worlds Collapse: What History, Systems, and Complexity Can Teach Us About Our Modern World and Fragile Future,"* 10.
[86] For a history of these distinctions, see "'It's Better to Forget Physics': The Idea of the Tactical Nuclear Weapon in the Early Cold War," *Physics in Perspective,* https://link.springer.com/article/10.1007/s00016-020-00251-3.

guide high-impact giving. Understanding this fact allows us to then make a conclusion; *all else equal*, it is more cost-effective if a given grant prevents large nuclear wars than if it prevents small nuclear wars.

The following section discusses the implications of this in greater detail; the highly non-linear relationship of war size to war costs, especially when considering long-term effects, only underscores the importance of focusing on the largest wars.[87]

The Non-Linearity of War Effects

A key feature of nuclear war is that the risk is non-linear. In large part due to climate dynamics, threshold effects, and the possibility of societal collapse, all-out nuclear wars are *disproportionately worse* than highly limited nuclear wars. We can demonstrate this nonlinearity by comparing the effects of the bomb used on Hiroshima in 1945 to the effects of 100 such bombs used in a hypothetical conflict between India and Pakistan. The best available estimates suggest that about 70,000 to 140,000 people died in Hiroshima alone.[88] Two such bombs likely would cause the deaths of 140,000 to 280,000 people. A naive linear extrapolation of this would suggest that 100 times as many bombs would lead to 100 times as many deaths — 7,000,000 to 14,000,000. As we will see, however, this linear extrapolation misses key effects that multiply the damage of larger nuclear wars. The next sections explain the various features of nuclear war that, although their magnitude is uncertain, stack together to create an approximately super-linear relationship between war size and war cost (though this relationship may break down at extreme war sizes, as discussed below in the section titled "Overkill and 'Making the Rubble Bounce'").

Nuclear Winter

The deaths from a war fought with 100 Hiroshima-sized weapons may be more than an order of magnitude higher than a naive linear extrapolation would suggest.This is in large part due to "nuclear winter" — the hypothesized climate effects that would result from soot being injected into the atmosphere during a nuclear exchange. This could block out sunlight, resulting in cooler temperatures, which would make it harder to grow food. The literature on nuclear winter is relatively small, and estimates of the severity of potential cooling effects vary widely. We have collected a table of several influential studies in Appendix 3. After some initial interest in nuclear winter in the 1980s, there was a lull in research, followed by a second wave of nuclear winter studies in the 2000s and 2010s. Several of the studies in this second wave, led by Professors Alan Robock and Brian Toon,

---

[87] Readers may object that we do not know how to keep small nuclear wars from becoming large nuclear wars. We discuss this objection in "Deep Uncertainty on Interventions" and in "What about the Nuclear Taboo?"
[88] https://thebulletin.org/2020/08/counting-the-dead-at-hiroshima-and-nagasaki/.

among others, focused on one hypothetical scenario: a war between India and Pakistan involving 100 15-kiloton warheads. This scenario is therefore among the best-studied.[89] Estimates of cooling in these studies range widely, from very limited cooling mostly in polar regions, to global mean cooling of nearly 2°C.[90] Xia *et al.* (2022) estimated that such a war would lead to 27 million direct deaths (from the impact of the bombs), and could lead to the starvation of over 250 million people due to climate effects.[91]

Reisner *et al.* (2018) provide one detailed critique of Robock *et al.*'s research. Many studies with high death toll estimates assume that human behavior will not adapt to new agricultural conditions.[92] Several studies also do not account for the interaction of climate change with nuclear winter — a 2°C cooling via soot injection in a 2°C-warming world could be treated analogously to geoengineering or "Solar Radiation Management" (SRM) via stratospheric aerosol injection, with effects on crop yields highly uncertain.[93] Some of the SRM literature suggests that volcanic cooling (analogous to nuclear war-induced cooling) may have only a small net negative impact on crop yields, or even a positive impact when accounting for climate change. Pongratz et al. estimated that SRM in a high (2X) CO_2 environment would actually *increase* global crop yield because it would decrease the stress of high temperatures while retaining CO_2 fertilization benefits.[94] A 2018 analysis in *Nature* suggested that mid-21st-century cooling of 0.88°C as a result of solar radiation

---

[89] Xia et al. "Global food insecurity and famine from reduced crop, marine fishery and livestock production due to climate disruption from nuclear war soot injection," *Nature Food;* Reisner et al., "Climate Impact of a Regional Nuclear Weapons Exchange: An Improved Assessment Based On Detailed Source Calculations" *JGR Atmospheres;* Pausata et al., "Climate effects of a hypothetical regional nuclear war: Sensitivity to emission duration and particle composition," *Earth's Future;* Mills et al., "Multidecadal global cooling and unprecedented ozone loss following a regional nuclear conflict," *Earth's Future;* Robock et al., "Climatic consequences of regional nuclear conflicts," *Atmospheric Chemistry and Physics*.

[90] Different assumptions about how much soot is injected, and how high into the atmosphere, drive much of this difference.

[91] Xia et al., "Global Food Insecurity and Famine from Reduced Crop, Marine Fishery and Livestock Production Due to Climate Disruption from Nuclear War Soot Injection," *Nature Food* 3, no. 8 (August 2022): 586–96, https://doi.org/10.1038/s43016-022-00573-0.

[92] "[W]e do not consider farm-management adaptations such as changes in cultivar selection, switching to more cold-tolerating crops or greenhouses and alternative food sources[...]." Xia *et al*. "Global Food Insecurity and Famine from Reduced Crop, Marine Fishery and Livestock Production Due to Climate Disruption from Nuclear War Soot Injection," *Nature Food* 3, no. 8 (August 2022): 586–96, https://doi.org/10.1038/s43016-022-00573-0.

[93] Thanks to Violet Buxton-Walsh for pointing out this similarity. Like nuclear winter, much of the science surrounding SRM is ultimately based on a limited understanding of explosive volcanic eruptions: "The closest natural analogues to these SRM proposals are major volcanic eruptions." Jonathan Proctor et al., "Estimating Global Agricultural Effects of Geoengineering Using Volcanic Eruptions," *Nature* 560, no. 7719 (August 2018): 480, https://doi.org/10.1038/s41586-018-0417-3.

[94] J. Pongratz et al., "Crop Yields in a Geoengineered Climate," *Nature Climate Change* 2, no. 2 (February 2012): 103, https://doi.org/10.1038/nclimate1373; for the degrees of SRM-induced cooling, see supplemental table S2.

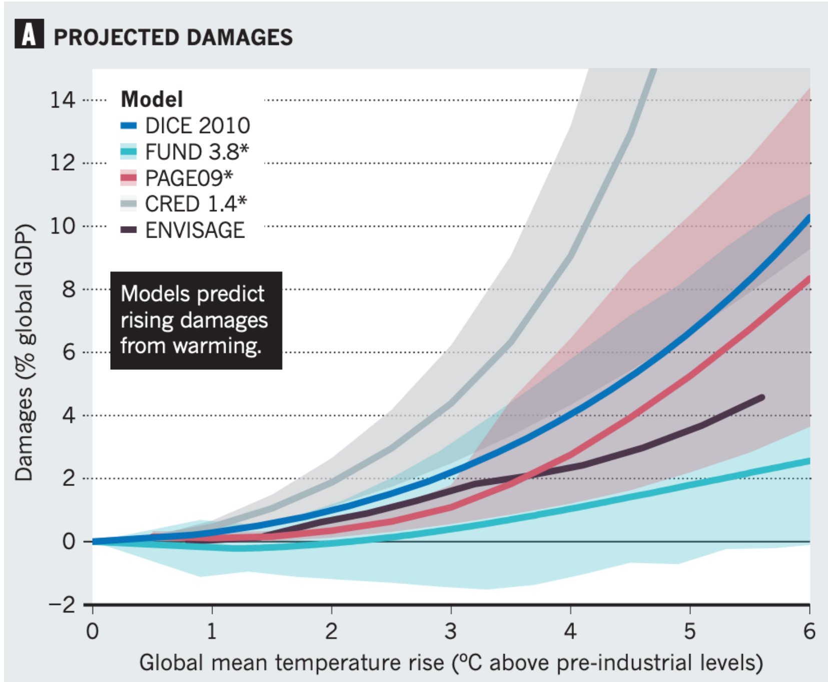
management would decrease heat stress on plants but have no discernible effect on maize, soy, rice, and wheat yields when accounting for the decrease in sunlight (i.e., damages from scattered sunlight and benefits from slight cooling may even out in some scenarios).[95] Nonetheless, we should not be sanguine about the possible negative effects of sudden cooling, such as frost disrupting normal crop growth.

However, even if a skeptic believed that these estimates were twice as high as they should be, they would be left with the possibility of more than 100 million people starving. This shows that deaths increase non-linearly. Again, 70,000 to 140,000 people died in Hiroshima alone; two such bombs likely would kill 140,000 to 280,000 people, but 100 times as many bombs would kill *more than* 100 times as many people. Reasoning in orders of magnitude, using the death tolls calculated in the nuclear winter literature suggests that moving from 1 to 2 15kt-sized nuclear explosions multiplies the deaths by a factor of 2, whereas moving from 1 to 100 15kt-sized nuclear explosions multiplies the deaths by a factor of 1,000.

Moreover, one implication of the possibility of nuclear war-induced global cooling is that there may be *threshold effects* roughly symmetrical to those of global warming. For instance, animals and crops domesticated during the Holocene epoch have certain temperature ranges, within which agriculture can thrive.[96] Crops can only withstand so many days of early frost at certain temperature floors (or days of extreme heat at certain temperature ceilings); as global climate change shifts the normal distribution of the number of such days, the mass of the distribution that passes these "thresholds" increases non-linearly.[97] Similar thresholds may exist for civilization as a whole — for example, a society may be able to handle only a certain amount of refugees before its institutions begin to crumble (e.g. due to xenophobic authoritarian backlash). Threshold effects are partly what drive the highly non-linear effects of climate change that are well-established in the literature on global warming:[98]

---

[95] Proctor et al., "Estimating Global Agricultural Effects," 482.
[96] Thanks to Matt Lerner for pointing me to this insight.
[97] Again, thanks to Matt Lerner for helping me understand this.
[98] Also cited in Ackva et al., *Changing Landscape,* 49.



Source: Richard L. Revesz et al., "Global Warming: Improve Economic Models of Climate Change," *Nature* 508, no. 7495 (April 2014): 173–75, https://doi.org/10.1038/508173a.

There are further uncertainties in the nuclear winter literature. One crucial consideration, for example, is how much of the black carbon that is produced from fires actually reaches the stratosphere; factors like time of year, cloud cover, and precipitation all affect this question.[99] Targeting plans also affect the amount of smoke produced, and it is generally agreed that military targets would produce less and different smoke than civilian city targets with higher fuel loads.[100] We discussed this targeting effect with Professors Robock and Toon, who agreed that targeting is a crucial consideration and that *in theory* targeting

[99] "A necessary condition for significant global cooling impacts from a limited exchange is that the atmosphere and fuel be conducive to fire and subsequent upward transport of BC. Unlike a volcano with superheated buoyantly driven exhaust material from the caldera moving rapidly upward and being rapidly transported into the stratosphere (Ogden, Wohletz, et al., 2008; Ogden, Glatzmaier, & Wohletz, 2008), environmental conditions can inhibit BC transport in nuclear exchange events. For example, if the exchange takes place during the winter, or when clouds are present, or if the winds are too high or too light to prevent fire spread, and/or if it is raining or has recently rained, the amount of BC produced and/or the amount that reaches the lower stratosphere could be relatively small." Jon Reisner et al., "Climate Impact of a Regional Nuclear Weapons Exchange: An Improved Assessment Based On Detailed Source Calculations," *Journal of Geophysical Research: Atmospheres* 123, no. 5 (2018): 2753, https://doi.org/10.1002/2017JD027331.

[100] "Likewise, if remote military sites are primarily targeted instead of cities, the amount of BC produced will also be substantially reduced." Jon Reisner et al., "Climate Impact of a Regional Nuclear Weapons Exchange: An Improved Assessment Based On Detailed Source Calculations," *Journal of Geophysical Research: Atmospheres* 123, no. 5 (2018): 2753, https://doi.org/10.1002/2017JD027331.

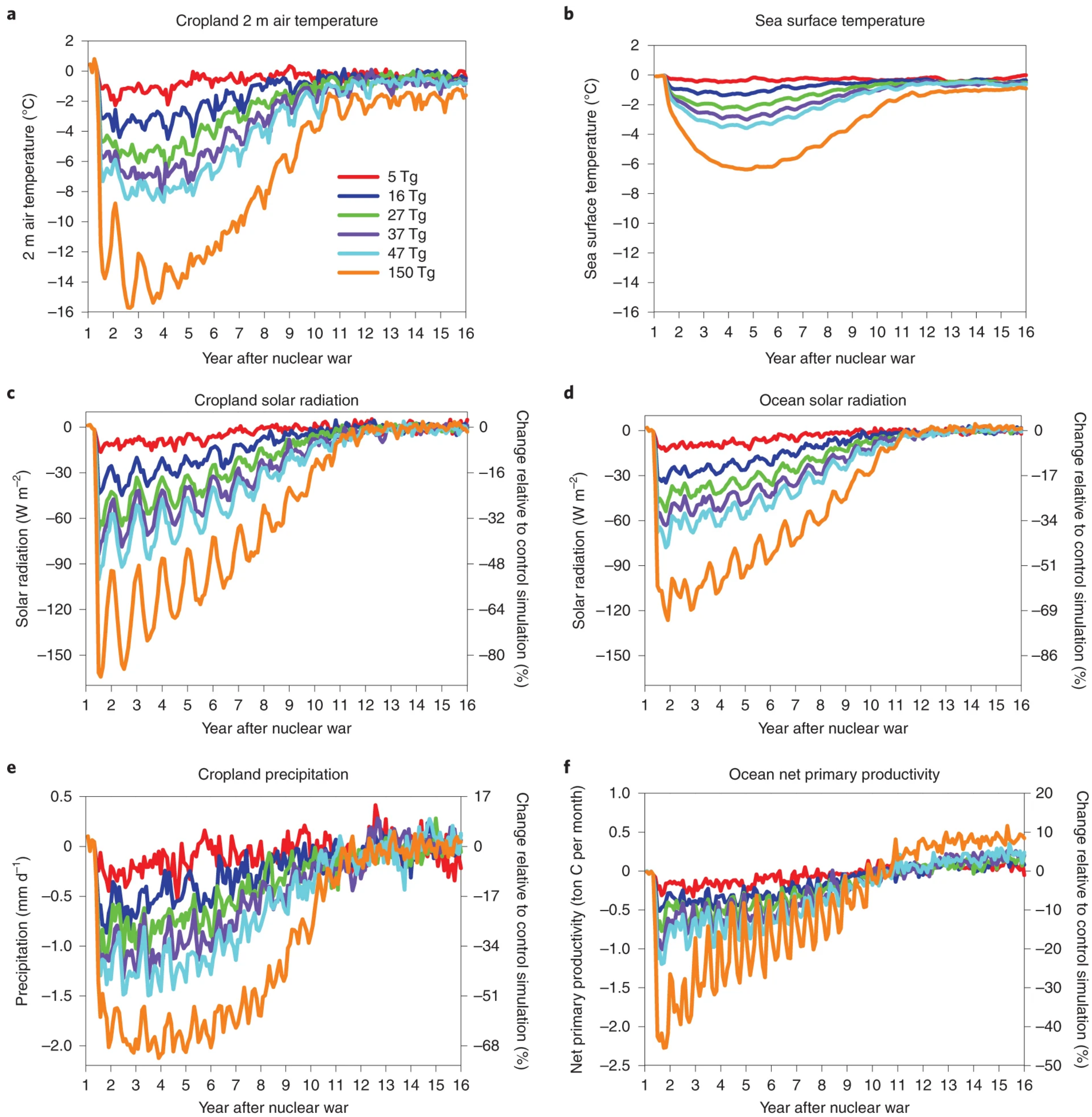
remote missile silos in desert-like areas (where China's missile fields are situated) would produce less smoke and therefore a less severe effect.[101] However, while publicly-stated U.S. targeting policy explicitly rejects civilian targeting, some military targets may necessarily be located close to civilian areas, and targeting policies may evolve in the course of a nuclear war, such that it is highly unclear *where* nuclear weapons would actually be used in a war.[102] Uncertainty on Russian and Chinese nuclear targeting is even larger.[103] Complicating the picture even further, the burning of massive fuel loads in a large nuclear war could release greenhouse gasses that lead to increased *warming* over longer time scales.[104]

In short, we simply don't know the answers to some key questions, including how much soot different kinds of nuclear war would release into the atmosphere and how much of this soot would reach the stratosphere. Nonetheless, there is some agreement that can be gleaned from the literature: very small nuclear wars would affect the climate much less than very large nuclear wars, and perhaps not at all.

This general pattern applies not just to temperature change, but to other important factors that affect agricultural production, as shown in figure 1. of Xia *et al.*'s 2022 *Nature Food* paper:[105]

---

[101] Professor Robock: "If you try and blow up a missile silo in the middle of nowhere, and you get a lot of dust, and there's nothing much to burn, it would have much less of an effect than when there are fires. But there are many, many military targets in cities." (Call with Professor Robock and Professor Toon, 2 March 2023)
[102] Under "Targeting," the *Report of the Nuclear Employment Strategy of the United States* states, "U.S. nuclear planning and targeting adhere to the laws of armed conflict. The United States has for decades rejected a deterrence strategy based on purposely threatening civilian populations, and the United States will not intentionally target civilian populations. The U.S. nuclear posture and alert status are tailored to enhance stability by ensuring that the United States retains sufficient survivable nuclear forces to ensure credible response options, no matter the nature of the adversary's attack." (*Report on the Nuclear Employment Strategy of the United States – 2020,* https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/NCB/21-F-0591_2020_Report_of_the_Nuclear_Employement_Strategy_of_the_United_States.pdf.)
[103] Thanks to Matthew Gentzel for this point.
[104] Thanks to Jon Reisner for pointing out this pathway.
[105] Lili Xia et al., "Global Food Insecurity and Famine from Reduced Crop, Marine Fishery and Livestock Production Due to Climate Disruption from Nuclear War Soot Injection," *Nature Food* 3, no. 8 (August 2022): 586–96, https://doi.org/10.1038/s43016-022-00573-0.



Source: "Climatic impacts by year after different nuclear war soot injections" in Xia *et al.,* "Global food insecurity and famine from reduced crop, marine fishery and livestock production due to climate disruption from nuclear war soot injection," *Nature Food.* "Tg." refers to teragrams of black carbon; again, the exact *type* of nuclear war that would inject such amounts of soot is disputed. Note that higher soot injection leads to greater climatic effects.

Therefore, while we don't know much about the probability distributions of nuclear winter, we can draw a stylized but useful conclusion from this literature: not all nuclear wars would lead to nuclear winter, but larger wars are more likely to lead to severe climate change, and threshold effects amplify the superlinear cost of this.[106] Once again, this may appear obvious, but it has large implications for the prioritization of marginal philanthropic resources.

Civilizational Collapse

Taking a longer-term perspective on humanity's future, the non-linearity of the war size-to-cost relationship becomes even more extreme. The starvation of 100 million people would be a humanitarian catastrophe and one of the worst events to befall humanity thus far, killing 1.25% of the world's population. (For comparison, between 17.4 million and 50 million people died in the 1918 "Spanish Flu" pandemic, one of the worst pandemics in recent history, representing between 0.95% and 5.4% of the population at the time.[107])

These are horrific numbers to contemplate. An even worse — and qualitatively different — picture emerges, however, when we consider the largest possible nuclear wars, such as an all-out U.S.-Russia nuclear exchange, and the possibility that such wars could lead to the collapse of global civilization.

As discussed above, experts disagree on the severity of global cooling caused by such a war. As far as we are aware, none of the scientists involved in nuclear winter modeling claim that direct extinction is likely even from severe nuclear winter. In the 1980s, some scientists suggested that a U.S.-Russia exchange (with more and bigger weapons than are available today) could trigger apocalyptic cooling that would bring land temperatures as low as -15°C to -25°C, worse than the coldest parts of the last Ice Age.[108] Today, no scholars of nuclear winter suggest such extreme effects, in part due to updated models, and in part due to reduced arsenals. Indeed, one common analogy for nuclear winter is the Toba Eruption approximately 74,000 years ago.[109] This is a common case study for solar radiation disruption and climate shocks, as it was the largest volcanic eruption in the Quaternary

[106] A crucial consideration is whether and how uncontrollably smaller nuclear wars may escalate into large nuclear wars. This is discussed in greater detail in the discussion of "right-of-boom philanthropy," below.
[107] Max Roser, "The Spanish Flu: The Global Impact of the Largest Influenza Pandemic in History," *Our World in Data*, accessed May 19, 2023, https://ourworldindata.org/spanish-flu-largest-influenza-pandemic-in-history.
[108] Turco et al., "Nuclear Winter: Global Consequences of Multiple Nuclear Explosions," *Science*.
[109] The rest of this paragraph is directly taken from an unpublished internal Founders Pledge shallow analysis by the author on the risks of large explosive volcanic eruptions.

period (the last 2.5 Million years).[110] The Toba "super-eruption" was at one point theorized to have caused a near-extinction event, but more recent estimates conclude that there was no volcanic winter-induced evolutionary bottleneck, and humanity was likely not close to extinction (note, however, that this eruption was before the Agricultural Revolution, and agricultural civilization may be vulnerable in different ways).[111] Although the absence of evidence is not evidence of absence, the very idea of "volcanic winter" was based on estimates from the 1990s that were too high by 1-2 orders of magnitude, according to a recent review of the evidence since then.[112] This matters because the idea of "nuclear winter" is analogous in some respects (though importantly dis-analogous in others) to "volcanic winter," and it should update our beliefs about the theory that either large volcanic eruptions or nuclear war could directly cause extinction events.[113]

Nonetheless, a U.S.-Russia nuclear war (or a future U.S.-China nuclear war with bigger arsenals) could in theory precipitate civilizational collapse; as discussed above, our uncertainty on nuclear winter effects ought to be very large, and this uncertainty goes both ways. The deaths of hundreds of millions to billions of people in such a cataclysm is an unprecedented event, and we do not know how civilization would respond, how robust and resilient the systems-of-systems underpinning global civilization are, and whether the survivors of such a catastrophe could rebuild. Modern civilization relies on a large human population and complex large-scale infrastructure. Because larger nuclear wars are more likely to lead to severe infrastructure damage and mass starvation than smaller nuclear wars, we can draw another simplified conclusion: not all nuclear wars would lead to civilizational collapse, but the largest wars will likely increase the probability of collapse.

Existential Catastrophe beyond Extinction

Scholars of nuclear winter agree that global cooling would probably not *directly* cause human extinction, and that there is high uncertainty around the estimates of the length and severity of the associated cooling. Again, even the scientists who have produced the most extreme estimates of nuclear winter effects do not claim that it would constitute an

---

[110] Clive Oppenheimer, "Limited Global Change Due to the Largest Known Quaternary Eruption, Toba ≈74kyr BP?," *Quaternary Science Reviews* 21, no. 14 (August 1, 2002): 1593–1609, https://doi.org/10.1016/S0277-3791(01)00154-8.
[111] Chad L. Yost et al., "Subdecadal Phytolith and Charcoal Records from Lake Malawi, East Africa Imply Minimal Effects on Human Evolution from the ~74 Ka Toba Supereruption," *Journal of Human Evolution* 116 (March 1, 2018): 90, https://doi.org/10.1016/j.jhevol.2017.11.005.
[112] Ibid.
[113] Some of this is addressed further below, under Making the Rubble Bounce. Important disanalogies include the proposed mechanism of lofting (massive forest fires may be better comparisons), the emission of sulfur dioxide by volcanic eruptions, and more. Thanks to Matthew Gentzel for pointing to this.

extinction event *per se* (except insofar as it might precipitate civilizational collapse).[114] A recent report by the Johns Hopkins Applied Physics Laboratory explains, "While some nuclear disarmament advocates promote the idea that nuclear winter is an extinction threat, and the general public is probably confused to the extent it is not disinterested, few scientists seem to consider it an extinction threat."[115] This was possibly even the case with the far larger arsenals at the height of the Cold War.[116]

The apparent low probability of *direct* extinction via nuclear winter has caused some analysts of existential risks to discount the threat of nuclear war.[117] The Existential Risk Observatory, for example, writes "complete extinction because of direct effects of nuclear war, although very important for non-existential reasons, is not the main existential threat."[118] A similar view is expressed in many other analyses, sometimes accompanied by low subjective point estimates of extinction risk.[119] These views, though diverse, are based on three types of consideration. First, considerations like the expiration dates of some critical resources lead some analysts to optimistic conclusions about post-collapse worlds; there may be a "grace period" where resources last long enough to support efforts to restart civilization.[120] Second, mechanistic views of a post-collapse society suggest quasi-Malthusian "collapse equilibria," where populations are balanced against available resources.[121] Third, statistical arguments about uncorrelated risks to disconnected groups

---

[114] The most detailed discussion of this can be found in footnote 36 on page 336 of Ord's *The Precipice*.
[115] James Scouras, "Nuclear War as a Global Catastrophic Risk," *Johns Hopkins Applied Physics Laboratory,* 3-4.
[116] See Martin, "Critique of Nuclear Extinction," *Journal of Peace Research*.
[117] For a discussion of the uncertainties surrounding civilizational collapse, see a recent chapter by Haydn Belfield on "Collapse, Recovery, and Existential Risk" in the 2023 volume *How Worlds Collapse: What History, Systems, and Complexity Can Teach Us About Our Modern World and Fragile Future*.
[118] ERO, "Nuclear War," Existential Risk Observatory, accessed February 20, 2023, https://www.existentialriskobservatory.org/nuclear-war/.
[119] E.g. Jeffrey Ladish, "Nuclear War Is Unlikely to Cause Human Extinction," accessed February 20, 2023, https://forum.effectivealtruism.org/posts/mxKwP2PFtg8ABwzug/nuclear-war-is-unlikely-to-cause-human-extinction. These seem to be based in part on a conflation of the words *extinction* and *existential.* E.g., Scouras, "But what nuclear war is not is an existential risk to the human race. There is simply no credible scenario in which humans do not survive to repopulate the earth." in "Nuclear War as a Global Catastrophic Risk," *Johns Hopkins Applied Physics Laboratory.*
[120] The term "grace period," coined by Lewis Dartnell in his popular book *The Knowledge* is one such consideration — during this period, it is argued, humanity will have the ability to live in remaining shelters, eat food in order of expiration date, and use or repurpose remaining fuel and raw materials (Dartnell, *The Knowledge: How to Rebuild Our World from Scratch.*)
[121] This is a quasi-Malthusian argument: fewer people need fewer resources, and the population level can be expected to adjust to the available resources. For example, Rodriguez writes, "I expect a large fraction of the survivors who survive this initial catastrophe to die during a period of violent competition following the collapse. I expect this period of violence to last until the survivors reach some sort of equilibrium, where they perceive there to be sufficient food, water, and shelter. The reason for this view is that it seems very irrational for groups to keep fighting if there are enough resources to go around." Rodriguez, "What is the likelihood that civilizational collapse would directly lead to human extinction?"

make total extinction appear unlikely.[122] These views appear to be widely held among scholars of existential risk.[123]

These arguments should not, however, lead philanthropists to completely dismiss the risk of long-term existential catastrophe emanating from global catastrophic nuclear risks. In a chapter on "Collapse, Recovery, and Existential Risk" in the 2023 volume *How Worlds Collapse*, Haydn Belfield of the Centre for the Study of Existential Risk at the University of Cambridge provides a critique of the "sanguine view" of collapse risks: "a collapse could destroy humanity's longterm potential. We cannot yet confidently rule out the prospect of permanent collapse or extinction — and we have good reasons to be concerned that a global civilization that recovered may have much worse prospects."[124] Broadly, Belfield argues that there is a surprising lack of interaction between collapse scholars and existential risk scholars and outlines three broad critiques of the sanguine view: historiographic biases, a neglect of deep uncertainties about collapse, and a neglect of non-extinction existential risk originating from civilizational collapse.

First, there appears to be a widespread belief that collapse would knock humanity back to an earlier "stage" of history, from which progress could begin anew.[125] In fact, the idea that humanity would simply revert to a hunter-gatherer lifestyle and go through "stages" of increasing complexity is simplistic and possibly unrealistic — like the evolution of biological systems, cultural evolution is not directed towards a goal.[126] In the worst case, Belfield writes, this means that modern global civilization may sit atop a fragile perch from which

---

[122] In short, if one expects there to be multiple groups of survivors, and assumes that the risk to them is uncorrelated, then even extremely high probabilities of extinction for each individual group lead to a low overall probability of extinction. (See, e.g. Rodriguez, "What is the likelihood that civilizational collapse would directly lead to human extinction?")

[123] "I think that even if civilization did collapse, it would likely recover" (Ord, *The Precipice*, cited in Belfield, "Collapse, Recovery, and Existential Risk"); "Although such a catastrophe [WMD use and major war] is, in my view, unlikely to lead to unrecovered civilisational collapse, it is difficult to be extremely confident that it won't" (MacAskill, *What We Owe the Future,* 142); "I think it's exceedingly unlikely [<0.0001] that humanity would go extinct (within ~a generation) as a direct result of a catastrophe that causes the deaths of 50% of the world's population [...] I think it's very unlikely [0.01 to 0.1] that humanity would go extinct as a direct result of a catastrophe that caused the deaths of 90% of the world's population [...] I think it's fairly unlikely [0.1 to 0.3] that humanity would go extinct as a direct result of a catastrophe that caused the deaths of 99.99% of people" (Rodriguez, "What is the likelihood that civilizational collapse would directly lead to human extinction?")

[124] Belfield, "Collapse, Recovery, and Existential Risk," 63.

[125] "Writing on the prospects of a contemporary collapse has sometimes fallen into these simplistic terms, imagining progress simply reversed with humanity being 'knocked back' to a 'previous stage' – especially that of 'hunter-gatherers' (Hanson, 2008; Mitchell & Chaudhury, 2020). Popular media (books and films such as T*he Road*) have encouraged this imaginary." (Belfield, "Collapse, Recovery, and Existential Risk," 64.)

[126] "[H]istorical collapses have rarely led to foraging. Foraging is a highly skilled life-style, and many areas are now unsuited to it due to human land-use. Societies would not necessarily need to, or be able to, transition from farming and industry." (Belfield, "Collapse, Recovery, and Existential Risk," 64.)

gradual recovery is impossible: "Humanity may have climbed high up a 'rungless ladder' — while historic societies fell only a small distance, modern societies may reach 'terminal velocity'."[127]

Second, there is simply deep uncertainty about the risk of collapse. For example, there are risks like weapons of mass destruction (e.g. nuclear or biological weapons) falling into the hands of bad actors, for which there are very few historical analogies (only the managed collapse of the Soviet Union — see Case Study on Leverage: Cooperative Threat Reduction later in this report).[128] Most fundamentally, however, a collapse of our society is truly unprecedented.[129] Belfield writes, "a collapse of a global, industrialised society is unprecedented. We cannot be sure that extinction would not follow. The probability of extinction is low, but not imperceptibly low — it is not a rounding error from 0%."[130]

Finally, extinction is only one of several possible post-collapse worlds that present an existential risk.[131] Large-scale nuclear war may lead to the collapse of global civilization (e.g. through mass starvation caused by nuclear winter), which, in turn, could either lead to civilizational recovery, or to one of three existential catastrophes. The first is human extinction (although some analysts find this unlikely for the reasons discussed above).[132] The second is unrecovered civilizational collapse, in which global society is permanently stuck in a bad state (e.g. because key knowledge and resources such as easily-accessible fossil fuels are lost). The third is civilizational recovery with negative values, such as a totalitarian dystopia.[133]

---

[127] Belfield, "Collapse, Recovery, and Existential Risk," 67-68. Kemp uses the same analogy: "Think of civilisation as a poorly-built ladder. As you climb, each step that you used falls away. A fall from a height of just a few rungs is fine. Yet the higher you climb, the larger the fall. Eventually, once you reach a sufficient height, any drop from the ladder is fatal." Luke Kemp, "Are We on the Road to Civilisation Collapse?," *BBC*, February 18, 2019, https://www.bbc.com/future/article/20190218-are-we-on-the-road-to-civilisation-collapse.
[128] "During the process of breakdown, weapons of mass destruction (WMD) such as nuclear weapons or bioweapons are more likely to be used by states, or to fall out of state control" (Belfield, "Collapse, Recovery, and Existential Risk," 67.)
[129] As far as we know (see discussion of "observer selection effects" below).
[130] Belfield, "Collapse, Recovery, and Existential Risk."
[131] Some definitions of existential risk include collapse *per se* (see, e.g. Nathan Alexander Sears, "Existential Security: Towards a Security Framework for the Survival of Humanity," *Global Policy* 11, no. 2 (April 2020): 255–66, https://doi.org/10.1111/1758-5899.12800.)
[132] Luisa Rodriguez, "What Is the Likelihood That Civilizational Collapse Would Directly Lead to Human Extinction (within Decades)?," *Effective Altruism Forum*, 2020, https://forum.effectivealtruism.org/posts/GsjmufaebreiaivF7/what-is-the-likelihood-that-civilizational-collapse-would.
[133] Again, this is outlined in greater detail in Haydn Belfield's chapter "Collapse, Recovery, and Existential Risk" in *Miguel A. Centeno,* Peter W. Callahan, Paul A. Larceny, and Thayer S. Patterson (eds), *How Worlds Collapse*.

Belfield argues that there is a wide range of possible recoveries, including the global spread of totalitarian regime types, which would have broad negative consequences for the future of humanity.[134] In this case, civilizational collapse with a wide range of possible outcomes could be especially bad for a number of reasons:

1. There may be an asymmetry between possible bad outcomes and good outcomes (e.g. if the "floor" on possible dystopias is much lower than the "ceiling" on possible utopias is high, or if there are more possible bad worlds than good worlds).
2. Our current outcome (general peace and widespread liberalism) may be above average for possible worlds and a "reboot" would therefore lead to a "regression to the mean" with likely worse outcomes.
3. Good outcomes may be less likely in a recovery than bad outcomes (e.g. because recovery favors militarized totalitarian regimes).

For these reasons and more, Belfield argues that "value is fragile, recovery is risky, and we should not 'reroll the 100-sided die'."[135]

Practically speaking, therefore, philanthropists should not be fanatical about prioritizing *extinction risks* only; we cannot dismiss the possibility that the upper end of global catastrophic nuclear risk could present many kinds of existential risks to human civilization. This uncertainty in turn may lead philanthropists to re-prioritize risks that have a higher likelihood of occurring but a presumed low conditional likelihood of *direct extinction* (such as nuclear war) over risks that have a lower likelihood of occurring but a presumed high conditional likelihood of direct extinction (such as risks from atomically precise manufacturing or "nanotechnology"). Similarly, Belfield writes, "[If] there is a substantial probability that collapse could destroy humanity's long-term potential, this should change one's view of catastrophic risks: climate change, nuclear war and broad biological risks should become more important."[136] For the purposes of this argument, the uncertainty on the outcomes of civilizational collapse, and the possibility of long-term existential catastrophe, further add to the non-linearity of costs from nuclear war.

What about the Nuclear Taboo?

One possible objection to claims about the non-linearity of war-effects is that *any* nuclear use, even of just one weapon, would breach the "nuclear taboo" and open the floodgates to further nuclear use (and other WMD use). If this were true, then we ought to model the move from conventional war to small nuclear war to be highly discontinuous in terms of

[134] See "A Range of Recoveries" in Belfield, "Collapse, Recovery, and Existential Risk."
[135] Belfield, "Collapse, Recovery, and Existential Risk," 77.
[136] Belfield, "Collapse, Recovery, and Existential Risk," 79.

expected cost, leading to a very different risk-reduction strategy. This section outlines reasons for being skeptical of this influential claim.

To summarize the basis for these claims, some scholarship in international security suggests that there are normative restraints — including "taboos" or strong non-use norms — on weapons of mass destruction.[137] These scholars argue that the historical record of presidential decision-making and rhetoric provides evidence for the existence of such a taboo.[138] Thus, if the taboo is broken, there are several concerns. A broken nuclear taboo may result in a renewed arms buildup and more frequent use of large nuclear weapons. As of recently, the United States will retire its megaton-range weapons (B83-1); nuclear weapons states appear to understand that extremely large nuclear warheads — like Tsar Bomba — are an inefficient use of fissionable material.[139] If states' view of megaton-range weapons were to change (e.g. for reasons of prestige or national glory, if not strategic utility), then ever larger stockpiles and larger deployed arsenals could perhaps pose an elevated risk in a way that current moderate stockpiles do not. Perhaps more concerningly, it could be the case that the breach of the WMD taboo via nuclear weapons use could also make biological weapons development and use more likely, because WMD as a whole are no longer considered off limits — which in turn could lead to civilizational collapse and human extinction.

For the sake of argument, in this section we assume that a nuclear taboo does exist. First, the historical evidence given by scholars of the taboo does suggest that decision-makers sometimes at least act *as if* they observed such a taboo. Second, we want to engage with the argument on its own terms: assuming that a norm against nuclear use plausibly exists, are there strong reasons to believe that a norm breach would likely lead to widespread use or escalation? It is unclear from the literature on international relations how much norms matter, but evidence from earlier attempts to ban or stigmatize certain weapons classes suggest that once such a norm is breached, norm weakening *can* follow, but so can norm strengthening.[140]

---

[137] Nina Tannenwald, "The Nuclear Taboo: The United States and the Normative Basis of Nuclear Non-Use," *International Organization* 53, no. 3 (1999): 433–68.
[138] Ibid.
[139] "The 2022 Nuclear Posture Review: Arms Control Subdued By Military Rivalry," *Federation Of American Scientists* (blog), accessed November 7, 2022, https://fas.org/blogs/security/2022/10/2022-nuclear-posture-review/. (Because bombs explode in three dimensional space, the destroyed area of the target on the surface of the earth only scales approximately by a cubic root, making multiple lower-yield weapons more effective by total weight in targeting a given area.)
[140] For an argument *against* the importance of international institutions and norms in shaping international security, see John J. Mearsheimer, "The False Promise of International Institutions," *International Security* 19, no. 3 (1994): 5, https://doi.org/10.2307/2539078.

Sometimes, norm breach appears to engender repeat use. For example, when would-be norm-violators believe that the costs of punishment for a violation are unlikely to outweigh the benefits of use, they may decide that the cost-benefit calculation justifies breaking an apparent norm.[141] For example, in the early twentieth century, nations around the world attempted to negotiate a ban on submarine warfare and aerial bombardment, but the restrictions were quickly breached with the onset of World War II.[142] The restrictions on submarine warfare did not survive the beginning of the war and were immediately broken. The restrictions on aerial bombardment were respected for several months in the early phases of the war, but once violated, resulted in large-scale bombing campaigns of civilian targets, culminating in the use of atomic bombs on Japan.[143] It should be noted that it is not clear from either of these cases whether there ever truly existed a "taboo," or whether the attempted "bans" were meaningless from the very beginning. That is, it is not clear whether these apparent examples are in the same reference class as the nuclear taboo. Both were legalistic "bans" on weapons that were broken, rather than normatively-constraining and deeply-held moral beliefs about right and wrong (as the nuclear taboo may involve, according to some).

Indeed, some scholars suggest that the breach of a norm against the use of weapons can strengthen that very norm; for example, when norm breach is punished. James Scouras has argued that occasional norm breach may in fact be necessary for a norm to remain in place: "norms, in general, cannot endure indefinitely without periodic violations that provide tangible reminders of their value" — and of the consequences of violation.[144] Similarly, in their discussion of North Korean nuclear first use, Hahn *et al.* write:

> "There is a presumption that once violated, the norm against the use of nuclear weapons cannot endure. But, this presumption is not based on a body of research; it is possible that the response to first use could act to reaffirm the relevance of the norm and that a single violation would not necessarily irreversibly undermine the norm's existence."[145]

---

[141] Thanks to Matthew Gentzel for his helpful advice on the importance of the consequences of norm violation in this section.
[142] Paul Scharre and Megan Lamberth, "Artificial Intelligence and Arms Control," accessed November 1, 2022, https://www.cnas.org/publications/reports/artificial-intelligence-and-arms-control.
[143] Jeffrey W. Legro, "Which Norms Matter? Revisiting the 'Failure' of Internationalism," *International Organization* 51, no. 1 (1997): 31–63.
[144] Scouras, "Nuclear War as a Global Catastrophic Risk," 8.
[145] Erin Hahn, "Responding to North Korean Nuclear First Use: So Many Imperatives, So Little Time," 9. (Also cited in Scouras, "Nuclear War as a Global Catastrophic Risk," 8.

Is there evidence for these claims? The literature on norm erosion and robustness is thin and data are sparse, but they suggest that we ought to once again be highly uncertain about the *sign* of the value of norm breach (i.e. whether its effects will be net positive or negative).[146] Strong claims that breaking the nuclear taboo *will* lead to further WMD norm erosion rather than norm strengthening are simply not supported by historical evidence. Rather, an important additional variable relates to the consequences of norm breach; when other states can impose harsh costs on the violator, they may be able to deter future violations.

One relevant recent example is the breach of the chemical weapons taboo via the use of these agents against civilians in Syria in the 2010s. Richard Price has argued that the evidence on balance suggests that the norm was *not* weakened by this violation.[147] First, the violation was met by strong third party condemnation, including limited conventional strikes in 2017 and 2018 by the United States and its allies that were *explicitly* framed as responding to the violation of the chemical weapons norm.[148] Second, Syria's violation did not lead to widespread use of chemical weapons by the 200+ other states in the international system, even after the initial use of chemical weapons was *not* punished strongly by the Obama administration.[149] Similarly, aggregated probabilistic forecasts on crowd forecasting sites now assign low probabilities to the use of chemical weapons in the Russo-Ukrainian war.[150] On balance, therefore, if we believe that nuclear use could be similar to chemical weapons use, it is not clear that isolated norm violation would necessarily weaken (rather than strengthen or not affect) the nuclear taboo.

Finally, as emphasized later in this report, *accidents happen,* and the historical frequency of "close calls" suggests that we ought not to be sanguine about the possibility of nuclear accidents. It is possible that miscalculation will lead to the first use of nuclear weapons, at which point the norm will have been breached. The world ought to be prepared for this scenario.

---

[146] For a recent overview, see Nicole Deitelhoff and Lisbeth Zimmermann, "Norms under Challenge: Unpacking the Dynamics of Norm Robustness," *Journal of Global Security Studies* 4, no. 1 (January 1, 2019): 2–17, https://doi.org/10.1093/jogss/ogy041.
[147] Richard Price, "Syria and the Chemical Weapons Taboo," *Journal of Global Security Studies* 4, no. 1 (January 1, 2019): 37–52, https://doi.org/10.1093/jogss/ogy040.
[148] Ibid., 45-46.
[149] Ibid., 47. Notably, there *has* been use of toxic agents for assassinations by North Korea and Russia, but this is arguably a different use case than that outlawed in war. (Ibid.)
[150] Specifically, <5% probability of *accusation* of use. Good Judgment, "Before 1 April 2023, Will a NATO Member State Accuse Russia of Using a Chemical or Biological Weapon in Ukraine?," Good Judgment Open, accessed February 20, 2023, https://www.gjopen.com/questions/2513-before-1-april-2023-will-a-nato-member-state-accuse-russia-of-using-a-chemical-or-biological-weapon-in-ukraine.

Uncertainty on Existential Risk Pathways

There are strong philosophical arguments for why even low probabilities of existential risks can have overwhelming moral importance.[151] Because of the high value of preventing existential risks, even low probabilities on pathways like civilizational collapse contribute to the non-linear nature of the war size to war cost curve.[152] These non-extinction existential risks emerge *at some point* on the escalation ladder.[153] Importantly, we do not know how adding or removing "rungs" from the escalation ladder shapes the probability of reaching the rungs at the end (existential catastrophe); each step can be a chance to reset or to spiral further.[154] In our assessment, such existential risks would likely emerge *somewhere* between a highly limited regional nuclear war (akin to a bad natural pandemic) and a future all-out U.S.-Russia-China war (an unprecedented disaster for global civilization), and the likelihood increases as war size increases. The exact point at which these risks become large is unclear — indeed, some risks may turn out to be overblown, as discussed above. As described below, however, we only need a relatively coarse distinction between minor nuclear wars and major nuclear wars to guide philanthropic decision-making in the near-term.

Putting it all together

The previous sub-sections explained three intuitions about nuclear war:

---

[151] Hilary Greaves and William MacAskill, "The Case for Strong Longtermism," Global Priorities Institute Working Papers (Oxford, UK, June 14, 2021), https://globalprioritiesinstitute.org/hilary-greaves-william-macaskill-the-case-for-strong-longtermism-2/ and Nicholas Beckstead, "On the Overwhelming Importance of Shaping the Far Future" (New Jersey, Rutgers University, 2013). For an argument based on more traditional policy analysis and cost-benefit calculations, see Jason G. Matheny, "Reducing the Risk of Human Extinction," *Risk Analysis* 27, no. 5 (2007): 1335–44, https://doi.org/10.1111/j.1539-6924.2007.00960.x.

[152] This nonlinearity was stated most clearly by the philosopher Derek Parfit in *Reasons and Persons*:

> "I believe that if we destroy mankind, as we now can, this outcome will be much worse than most people think. Compare three outcomes:
>
> "1. Peace
> "2. A nuclear war that kills 99% of the world's existing population.
> "3. A nuclear war that kills 100%
>
> "2 would be worse than 1, and 3 would be worse than 2. Which is the greater of these two differences? Most people believe that the greater difference is between 1 and 2. I believe that the difference between 2 and 3 is very much greater… If we do not destroy mankind, these thousand years may be only a tiny fraction of the whole of civilized human history."

(Derek Parfit, *Reasons and Persons*, (Oxford: Clarendon Press, 1987), 453.) Similar reasoning applies to non-extinction existential risks.

[153] "Non-extinction existential risks", in this case, refers to risks that could permanently curtail the potential of human civilization over the long term.

[154] Thanks to Matthew Gentzel for this point.

1. Not all nuclear wars would lead to nuclear winter, but larger wars are more likely to lead to more severe climate effects.
2. Not all nuclear winters would lead to civilizational collapse, but more severe cooling will likely increase the probability of collapse.
3. Not all civilizational collapse events are necessarily existential catastrophes like unrecovered collapse, but more severe wars are more likely to lead to existential catastrophe.

Overkill and "Making the Rubble Bounce"

One key question for further investigation is whether the superlinearity of war effects levels off at certain war sizes. For example, nuclear winter effects may be self-limiting at a certain point, due to the physics of soot particles.[155] Scientific understanding of the stratospheric chemistry of large explosive volcanic eruptions has advanced in recent years, showing that the large amounts of sulfur aerosols involved in high-magnitude eruptions collide and form larger particles that fall out of the atmosphere before they reach the stratosphere and can cause longer-term climate effects.[156] This explanation is backed by the work of Claudia Timmreck *et al*., which suggests that "A huge atmospheric concentration of sulfate causes higher collision rates, larger particle sizes, and rapid fall out, which in turn greatly affects radiative feedbacks."[157] Relatedly, once a large amount of soot is in the stratosphere, the marginal effect of each additional soot particle decreases, because so much sunlight is already blocked.[158] Finally, "overkill" — assigning multiple warheads to single targets — and the simple problem of running out of useful targets — could level off the non-linear effects for extremely large wars (although it is possible that additional energy could make particles more likely to be lifted into the stratosphere)..[159]

At the very far end of possible war sizes, it is true that the largest wars imaginable — with unprecedented arsenals bombing all liveable areas over and over again — would flatten the curve of expected costs. In Winston Churchill's memorable phrase, at some point the parties of the war are simply "mak[ing] the rubble bounce."[160] At this point, the shape of war costs may follow a sigmoid curve, with superlinear effects leveling off *at some point*.

---

[155] Jon Reisner et al., "Climate Impact of a Regional Nuclear Weapons Exchange: An Improved Assessment Based On Detailed Source Calculations," *Journal of Geophysical Research: Atmospheres* 123, no. 5 (2018): 2767, https://doi.org/10.1002/2017JD027331.
[156] Thanks to Dr. Michael Cassidy for pointing me to this effect.
[157] Claudia Timmreck et al., "Aerosol Size Confines Climate Response to Volcanic Super-Eruptions:," *Geophysical Research Letters* 37, no. 24 (December 2010): 1, https://doi.org/10.1029/2010GL045464.
[158] Thanks to Professor Toon for pointing out this effect.
[159] Thanks to Professor Toon for this point (call with Professor Robock and Professor Toon, 3 March 2023), and to Matthew Gentzel for pointing to the uncertainty of adding additional energy.
[160] James Reston, "'Why Make The Rubble Bounce?,'" *The New York Times*, March 31, 1976, sec. Archives, https://www.nytimes.com/1976/03/31/archives/why-make-the-rubble-bounce.html.

We hope that future research can help to resolve the questions of *where exactly* superlinear cost flattens into rubble-bouncing, if it does. For the purposes of philanthropic prioritization, however, we do not need to know *exact* answers to these questions, but can use the structure of the problem to guide decision-making under uncertainty. In fact, the *exact* shape of the yield-to-cost curve may be complicated, with bumps and troughs shaped by targeting strategy, the locations of military targets, weapon type, and other considerations.[161] Moreover, the size of arsenals does not neatly translate into warhead size. For example, larger arsenals may be more able to sustain protracted nuclear war, a world of repeated nuclear use over extended periods of time — a scenario which has received almost no scholarly attention.[162] Such a world, with repeat large-scale nuclear use over several years, may suffer a prolonged nuclear winter with far more severe consequences than expected by current models, because the soot that left the stratosphere would be replaced via new nuclear use.

We ought not to get too bogged down in this. Instead, we can step back to appreciate a simple guiding insight about the cost of large all-out wars that remains robust to local variations. Because the field of post-Cold War nuclear war security studies has barely begun to reason about scenarios to the "right of boom" (see below), the option space of possible interventions to fund remains small. This means that, in practice, we are not prioritizing on a smooth or continuous spectrum of interventions that try to focus on specific war sizes. Rather, all we need for philanthropic prioritization is the coarse insight that "big wars as a category are disproportionately worse than single nuclear use or highly limited nuclear wars," even if we are uncertain about the exact shape of the curve.

Again, amidst all this uncertainty, this section suggests that philanthropists can draw one simple conclusion: the relationship of war size to expected cost is roughly super-linear, such that large nuclear wars are disproportionately worse than single nuclear use or highly limited regional nuclear exchanges.

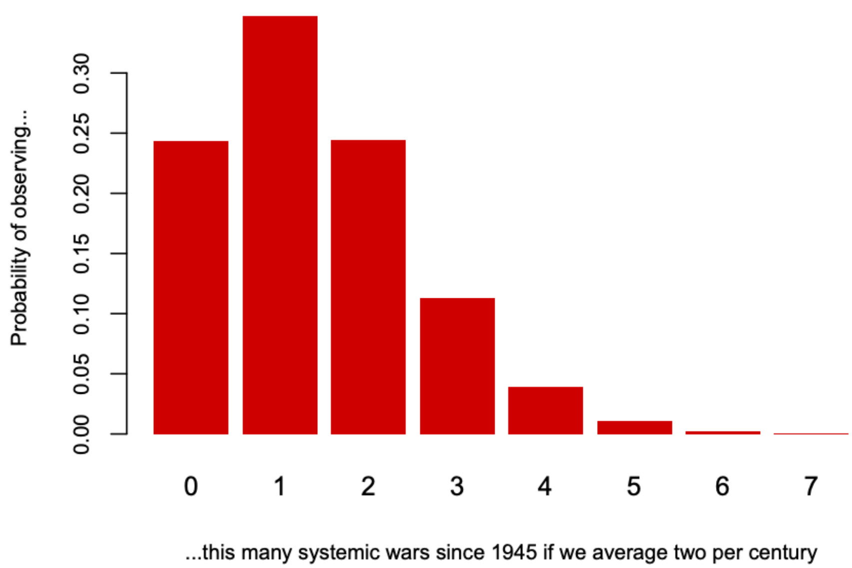
Deep Uncertainty on Interventions

The next key fact about the structure of the problem, as discussed in the introduction, is that there exists deep uncertainty about the effects of possible interventions. At the most extreme end of this uncertainty, it's *possible* that the current configuration of nuclear

[161] For example, a weak nuclear weapon state (or even a non-state actor) with a small arsenal may adopt a policy of targeting civilian population centers rather than military installations to maximize the horror and deterrence of their threats.
[162] Thanks to Dr. James Scouras for pointing me to this issue.

capabilities is among the most stable — and is indeed what has helped humanity survive the nuclear age so far. Nuclear weapons clearly have had deterrent power in *some* situations; for example, Nikita Kruschev explicitly told his top advisors in 1962: "In their time, the USA did the same thing [stationing missiles near the adversary], having encircled our country with missile bases. This deterred us."[163] Some go further, however, and argue that nuclear weapons have, for over three-quarters of a century, prevented another great power war that might have killed millions upon millions of humans.[164]

There are two reasons to be skeptical of such extreme claims. The first is that the apparent "Long Peace" is not statistically as surprising as it might look when eyeballing the apparent recent decline of warfare. As the political scientist Bear Braumoeller illustrates at length in his book *Only the Dead,* for example, observing zero systemic wars since 1945 is not all that surprising if one assumes an average of two great power wars per century:[165]



[163] Sergey Radchenko and Vladislav Zubok, "Blundering on the Brink: The Secret History and Unlearned Lessons of the Cuban Missile Crisis," *Foreign Affairs.*
[164] James Scouras, "Nuclear War as a Global Catastrophic Risk," *Journal of Cost-Benefit Analysis,* 275.
[165] Figure 2.3 "The entirely probable long peace" in Braumoeller, *Only the Dead,* 27. Notably, the assumption of "two great power wars per century" depends on how one codes "great power war"; Braumoeller argues that "if we look at the record of Great Power wars over the past five centuries (Levy, 1983; Goldstein, 1988, 146), we find an average of about two per century prior to the twentieth century" (26), citing Jack Levy's *War in the Modern Great Power System* and Joshua Goldstein's *Long Cycles: Prosperity and War in the Modern Age*. The point here is to illustrate what probabilistic conclusions one can draw based on one plausible assumption, rather than to defend the assumptions.

Source: Braumoeller, *Only the Dead: The Persistence of War in the Modern Age,* 27.

Similarly, as discussed in the next section, the prevalence of nuclear close calls since 1945 suggests that the world is perhaps not as safe as proponents of the Nuclear Long Peace hypothesis suggest.[166]

More broadly, debates about the probability of war may miss a bigger point about the overall risk, which again is a function of probability and consequences.[167] It is possible that nuclear weapons have decreased the probability of war while increasing its costs were it to occur. Indeed, this may continue longer historical trends in great power warfare: diminishing frequency and increasing costs.[168] If this is the case, the net impact of a decline in the probability of war on war's expected cost — if any — remains unclear. To the extent that costs may increase non-linearly the larger a war gets for some war sizes (discussed throughout this report), a decrease in war probability along with an increase in war size may lead only to *greater* catastrophes.

In short, philanthropists focused on nuclear security ought to acknowledge that the evidence is complex and uncertainty is high in this debate; we do not know how much, if at all, nuclear weapons contribute to international stability on net. This uncertainty suggests starting mostly with uninformed priors across different interventions. It suggests, for example, that the relative neglectedness of interventions can guide philanthropists towards areas where we ought to expect low-hanging high-impact fruit (see Philanthropy to the Right of Boom).

Is Disarmament Obviously Net-Positive?

We ought to stress-test these ideas thoroughly, given the importance of the topic. This section seeks to stress-test the claim that there is extreme uncertainty about the *sign* of interventions using one of the strongest counterarguments — wouldn't total nuclear disarmament be obviously net positive (were it politically feasible)?

One way that nuclear disarmament could be net-negative is if the benefits of the apparent Long Peace outweighed the costs of nuclear weapons. In this section, however, we put this question aside — as discussed earlier, the existence of the Long Peace would not be statistically surprising in a world where the underlying probability of war had not changed.

[166] Unless most instances of close calls are exaggerated or were not "close," and thus would not have resulted in nuclear war.
[167] Thanks to Matthew Gentzel for pointing to the framing that inspired this paragraph in a round of external reviews.
[168] See, e.g. Jack S. Levy, "Historical Trends in Great Power War, 1495-1975," International Studies Quarterly 26, no. 2 (1982): 278-300.

There exists a worse problem, however; if states fundamentally seek strategic weapons capabilities, might they develop a "substitute deterrent" that was even more dangerous than nuclear weapons? The answer appears to be yes.

The history of natural pandemics shows that biological agents — like plague and influenza — can cause human death tolls on a par with those of nuclear wars. Evolution, however, optimizes for reproductive fitness, not lethality, suggesting that pathogenic agents artificially engineered for maximum destructiveness could create even greater damage, potentially up to and including human extinction.[169] Theft from weapons labs, moreover, may be far worse with biological weapons than with nuclear weapons. A single stolen nuclear weapon or moderate amounts of missing fissile material would be great cause for concern, but difficult to deliver, and limited in effect. A single stolen sample of a weaponized infectious disease agent, on the other hand, could infect the entire world.[170] The same logic applies to unintentional release; accidents at nuclear weapons facilities are locally limited, whereas laboratory leaks from high-containment biology laboratories could again infect the entire world. Finally, the problems of mutation and self-replication are nearly unique to biological weapons,[171] and make the loss of human control a large problem for biological weapons, even those weapons initially designed to be limited in scope.[172] All else equal, therefore, a world with high bioweapons activity could be significantly worse than a world with high nuclear activity. Humanity and those working on catastrophic risk reduction should therefore be extremely concerned about the risk that states might substitute bioweapons for nuclear weapons, even if there is a low probability of substitution.

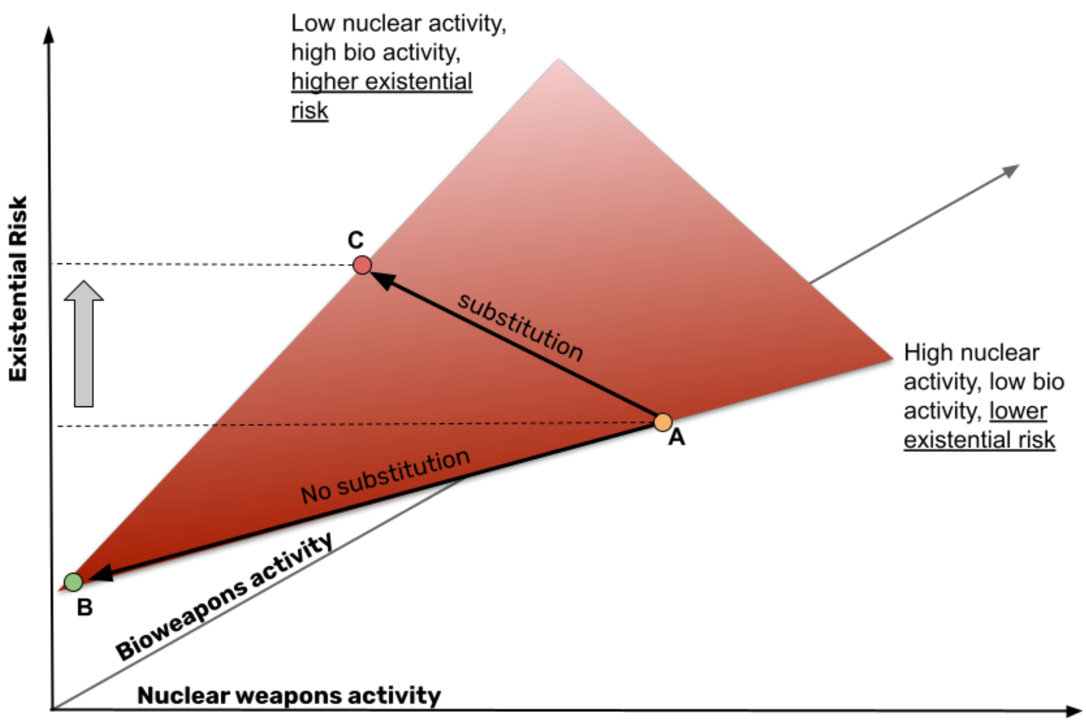
Borrowing from evolutionary biology, we can think of this as an existential risk "fitness landscape" with both a global optimum and local optima, as illustrated in the following figure. We can imagine an x-axis of "biological weapons activity," a y-axis of "nuclear weapons activity", and a z-axis of "existential risk." The global optimum is a world without nuclear weapons and without biological weapons. We may currently be at a local optimum — a world with high nuclear weapons activity, but low biological weapons activity. It may be

[169] Ibid., 135.
[170] For one such scenario, see Kevin Esvelt, "Delay, Detect, Defend: Preparing for a Future in Which Thousands Can Release New Pandemics," *Geneva Centre for Security Policy* accessed May 19, 2023, https://www.gcsp.ch/publications/delay-detect-defend-preparing-future-which-thousands-can-release-new-pandemics. In practice, misuse is still not as simple as sometimes implied, but the basic threat picture remains. For an overview of these considerations, see Sonia Ben Ouagrham-Gormley, *Barriers to Bioweapons: The Challenges of Expertise and Organization for Weapons Development.*
[171] They may, in different ways, also apply to offensive cyber operations, where the possible loss of control can give these tools an analogous "viral" character, as demonstrated by the Stuxnet virus, and possibly to the unsafe deployment of future AI systems.
[172] For a discussion of the problem of loss of human control, see Richard Danzig, "Technology Roulette: Managing Loss of Control as Many Militaries Pursue Technological Superiority" (Center for a New American Security, 2018), https://www.cnas.org/publications/reports/technology-roulette.

the case that in attempting to move from the local to the global optimum, the world slides into a worse state than before — a world with high biological weapons activity and low nuclear weapons activity:

The Existential Risk Landscape and Weapons Substitution

Substitution (moving from world A to world C) increases total existential risk even as nuclear risk decreases. The key question is how likely A→C is and how likely A→B is.



Source: Author's diagram.

Such substitution may occur for a variety of reasons. For instance, post-war nuclear disarmament may lead states to seek other strategic deterrents, and states appear to have historically viewed biological weapons programs as similar in kind to nuclear weapons programs. According to the bioweapons scholar Malcolm Dando, part of the reason for the Nixon administration's abandonment in the late 1960s and early 1970s of biological weapons was not only that biological weapons were less useful than nuclear weapons, but that they were more difficult to monopolize:

> "[T]he increasing realization that very few nations could acquire nuclear weapons of mass destruction, because of the difficulty of producing the fissile material, whereas many states could develop biological agents of mass destruction. Why not then renounce biological weapons, spread the idea that they were of little use, and even negotiate a total ban on them?"[173]

Similar attitudes may have existed in France and Britain; interest in bioweapons declined as capabilities in nuclear weapons increased, as evidenced in a shift in budgetary allocations from the 1950s to the 1970s away from biological and towards nuclear programs.[174] The claim that biological weapons programs were seen as truly "on a par" with nuclear weapons programs is likely too strong. If this were the case, we would expect at least some variation among the post-war victors' choices of weapon. Instead, we see the same pattern — nuclear-for-bio substitution — playing out several times (though probably not independently, given the influence of alliance dynamics like NATO). Moreover, some government assessments at the time, like a Presidential Science Advisory Committee report, did support the view that biological weapons were militarily challenging (though other reports concluded the opposite).[175]

Biological weapons simply may have been less useful historically without sophisticated tailoring or strong and costly domestic biodefense programs, but advances in synthetic biology and applications of artificial intelligence may obviate some of these issues. This report intentionally avoids discussing specific bottlenecks in bioweapons development and use — and in the emerging technologies that may address these bottlenecks — to avoid spreading dangerous ideas. Stated generally, if powerful states were to pivot the talent and resources behind the nuclear weapons complex towards biological weapons, worrisome breakthroughs could occur. Even without such breakthroughs, the very fact of increased biodefense activities in high-containment laboratories could increase the surface area for accidents via laboratory leaks, increasing the risk of dangerous pathogens (possibly modified for increased virulence) escaping and causing devastating global pandemics.

Nonetheless, it does appear from this evidence that biological weapons programs can fill at least a *similar* "niche" to nuclear weapons. As Susan Martin has written, part of the UK and

---

[173] "Biological Warfare, 1945-1972" in Dando, *Bioterror and Biowarfare*. In the same vein, Jonathan Tucker has written that for the Nixon administration, "[s]ending the message that biological warfare was ineffective would help to discourage hostile nations from acquiring a 'poor man's atomic bomb' that could serve as a military equalizer" (see Tucker, "A Farewell to Germs," 128).
[174] "[B]iological weapons were viewed as on a par with nuclear weapons at the end of the second world war. Only when the UK obtained its own nuclear systems did interest in biological weapons decline." Ibid.
[175] Jonathan B. Tucker, "A Farewell to Germs: The U.S. Renunciation of Biological and Toxin Warfare, 1969–70," *International Security* 27, no. 1 (July 2002): 119, https://doi.org/10.1162/016228802320231244.

US decision to renounce biological weapons was that they "judged it unwise to develop weapons that might offer a relatively cheap and easy alternative to nuclear weapons."[176] In short, it is possible that even the aftermath of nuclear disarmament could be net-negative depending on states' response to disarmament — our uncertainties remain high even on interventions that seem obviously good at first glance.

Accidents Happen

The history of nuclear weapons illustrates one frightening fact — accidents happen. The Future of Life Institute has compiled a "Timeline of Close Calls"; a list of moments in history when the world came exceptionally close to nuclear use. Reports of *some* of these close calls are probably exaggerated. For example, the Able Archer exercises of November 1983 are sometimes held up as examples of a time when reckless behavior nearly drove the world to nuclear war, but recent historical scholarship suggests that the situation was far less risky.[177] Nonetheless, it is clear from many other examples that the world has often come very close to nuclear war.[178] For example, during the Cuban Missile Crisis, a sole officer — Vasili Arkhipov — vetoed a captain of a Soviet nuclear-armed submarine who wanted to launch an attack in response to non-lethal depth charges dropped by U.S. ships; had Arkhipov not been on that particular submarine, the captain may have had the votes to be able to launch a nuclear-armed torpedo, and potentially start a global nuclear war.[179]

It is not obvious how to update on the frequency of close calls. On the one hand, they may suggest that we live in an extraordinarily risky world. On the other hand, they may suggest that the nuclear threshold is difficult to cross, even when we come very close to it. A third option is that few of the so-called close calls were truly close and that few of the examples would truly have caused nuclear war given other existing checks in place.[180] Given the impossibility of examining counterfactual worlds, however, the *apparent* prevalence of close calls may make the world vulnerable to changes in how nuclear decisions are made. For example, as discussed in Founders Pledge's report on *Autonomous Weapons and*

---

[176] Susan B. Martin, "The Continuing Value of Nuclear Weapons: A Structural Realist Analysis," *Contemporary Security Policy* 34, no. 1 (April 2013): 182, https://doi.org/10.1080/13523260.2013.771042.
[177] Simon Miles, "The War Scare That Wasn't: Able Archer 83 and the Myths of the Second Cold War," *Journal of Cold War Studies* 22, no. 3 (2020), https://direct.mit.edu/jcws/article-abstract/22/3/86/95296/The-War-Scare-That-Wasn-t-Able-Archer-83-and-the?redirectedFrom=fulltext.
[178] "Accidental Nuclear War: A Timeline of Close Calls," *Future of Life Institute*, accessed January 6, 2023, https://futureoflife.org/resource/nuclear-close-calls-a-timeline/.
[179] Bryan Walsh, "60 Years Ago Today, This Man Stopped the Cuban Missile Crisis from Going Nuclear," Vox, October 27, 2022, https://www.vox.com/future-perfect/2022/10/27/23426482/cuban-missile-crisis-basilica-arkhipov-nuclear-war.
[180] Thanks to Matthew Gentzel for pointing to this third option in a round of external reviews.

Military AI, the integration of machine learning systems into nuclear command, control, and communications (NC3)-related systems and decision-support systems may create dynamics (such as automation bias) that push close calls "over the edge."[181]

Moreover, given the problem of selective de-classification, we may expect the rate of accidents and close calls to be even higher than the known cases suggest. In short, accidents happen, especially when low probabilities compound over long time periods. This means that *total prevention* of nuclear war is not a plausible goal of risk reduction over the long term. Rather, we ought to insure against the worst possible worlds where close calls turn into instances of nuclear use; if war breaks out, we ought to attempt to limit its consequences.

What Went Down Can Come Back Up

Arms control experts have pointed to the apparent collapse of the extended New START treaty as one of the most concerning developments in nuclear security — the agreement currently places limits on all deployed strategic weapons of the U.S. and Russia, and is a cornerstone of contemporary arms control.[182] As mentioned above, Russian president Vladimir Putin recently announced that Russia would suspend its participation in the New START arms control agreement, which had been signed in 2010 with the aim of reducing the number of deployed strategic nuclear weapons and delivery systems and establishing a new verification regime, including on-site inspections.[183] These developments illustrate that U.S.-Russia arms control gains can be undone. Indeed, recent analysis by the Federation of American Scientists suggests that the U.S. and Russia could very rapidly *double* the number of deployed warheads without new START:

[181] For a deeper overview of issues with AI in NC3 contexts, see Alexa Wehsener et al., "AI-NC3 Integration in an Adversarial Context," *Institute for Security and Technology*, 2023, https://securityandtechnology.org/wp-content/uploads/2023/02/AI-NC3-Integration-in-an-Adversarial-Context.pdf.
[182] James M. Acton, Thomas MacDonald, and Pranay Vaddi, "Reimagining Nuclear Arms Control: A Comprehensive Approach" (Carnegie Endowment for International Peace), accessed February 20, 2023, https://carnegieendowment.org/2021/12/16/reimagining-nuclear-arms-control-comprehensive-approach-pub-85938.
[183] Heather Williams, "Russia Suspends New START and Increases Nuclear Risks," *Center for Strategic and International Studies*, February 23, 2023, https://www.csis.org/analysis/russia-suspends-new-start-and-increases-nuclear-risks.

Estimates of Russian Strategic Forces

Currently deployed under New START

	Launchers	Warheads
ICBMs	321	834
SLBMs	160	640
Bombers	52	200*
TOTAL	**533**	**1674**

Possible upload without follow-on treaty

	Launchers	Warheads
ICBMs	321	1197
SLBMs	160	832
Bombers	52	600
TOTAL	**533**	**2629**

*Deployed warheads attributed to bombers are higher than those counted under New START because the treaty artificially counts only one warhead per deployed bomber. This number instead counts warheads stored at bomber bases.

Source: Matt Kora, "If Arms Control Collapses, US And Russian Strategic Nuclear Arsenals Could Double In Size," Federation of American Scientists.

China, moreover, has traditionally been reluctant to engage in arms control due to what it views as the imbalance between its nuclear forces and those of Russia and the United States.[184]

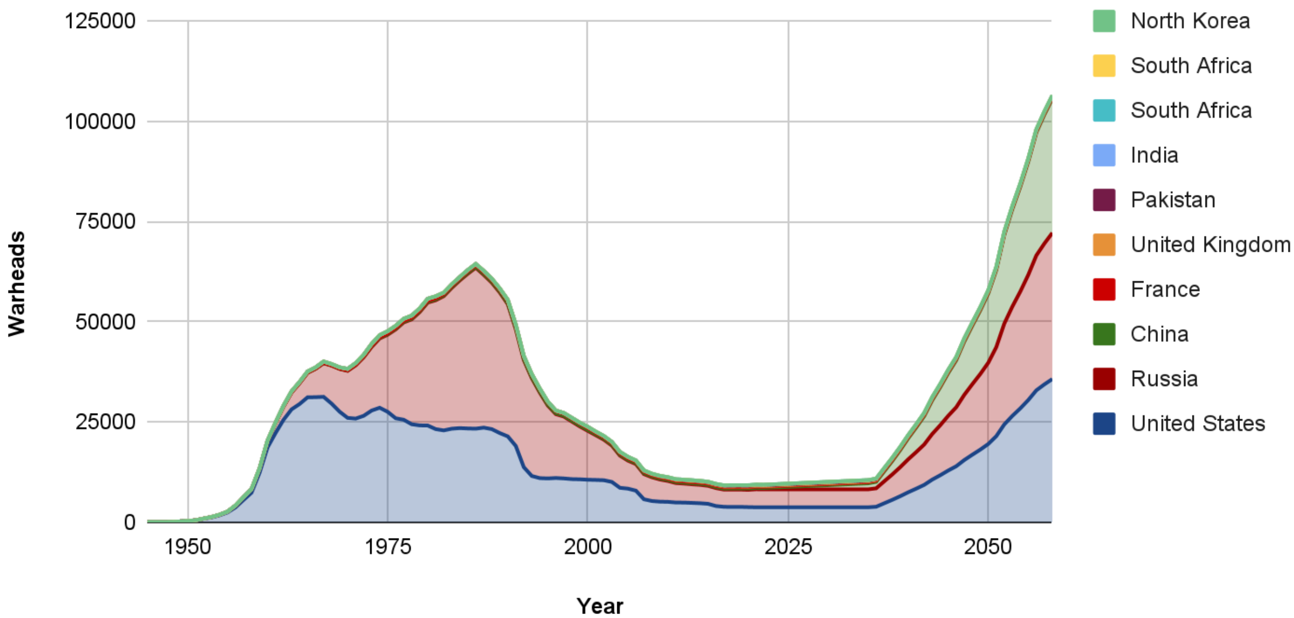
These facts portend bad possible futures, in which the arms control gains made in the last four decades could be undone — what went down can come back up. Imagine, for example, a world in which China, starting in 2035, chooses to arm in the same way that Russia armed between 1950 and 1985, that the United States matches this buildup weapon-for-weapon, and that Russia in turn matches the U.S. buildup weapon-for-weapon.[185] If the Soviet Union armed like this, then China can too. Indeed, one recent (2023) analysis of China's nuclear ambitions suggested that China's longer-term expansion of its nuclear capabilities may include "seeking to build an arsenal on par with Washington's and Moscow's."[186]

---

[184] See, e.g. "Beyond (Traditional) Bilateralism: A U.S.-Chinese Fissile Material Management Regime" in James M. Acton, Thomas MacDonald, and Pranay Vaddi, "Reimagining Nuclear Arms Control: A Comprehensive Approach" (Carnegie Endowment for International Peace), accessed February 20, 2023, https://carnegieendowment.org/2021/12/16/reimagining-nuclear-arms-control-comprehensive-approach-pub-85938.
[185] This may be strategically irrational, but so was, arguably, the Cold War arms race — strategic considerations are only one of many reasons that states choose to arm.
[186] "Some part of Beijing's buildup surely is meant to bolster its second-strike retaliatory capability in the face of what China perceives as shifts in U.S. conventional and nuclear capabilities and policies. China's long-term goal for the expansion, however, could be more ambitious and potentially even include seeking to build an

A Tripolar Arms Race

Combining OWID Estimates (1945-2022) with Cold-War-style arms-racing beginning in 2035.

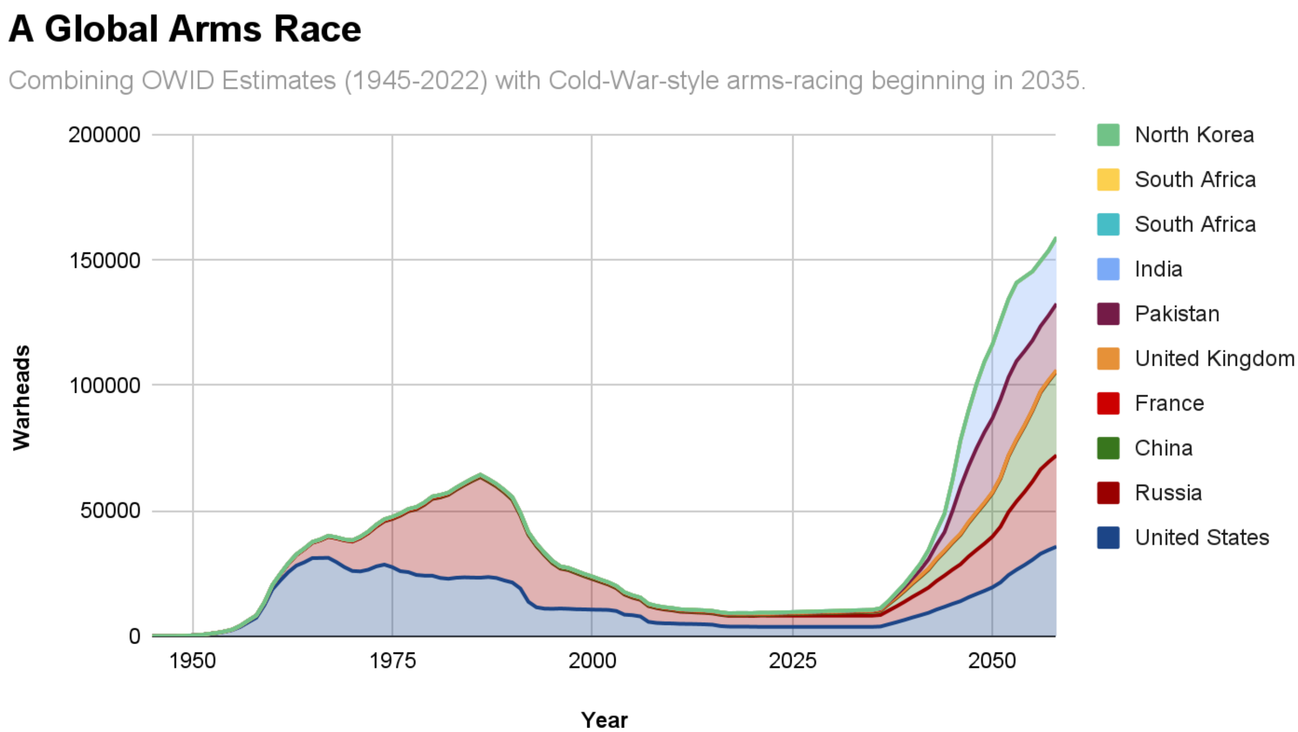


Source: Author's visualization based on *Our World in Data* dataset, assuming that China arms post-2035 as Russia did between 1960 and 1982, and that the U.S. and Russia match this build-up weapon-for-weapon.[187]

We can imagine, furthermore, that a rising India, threatened by the armament of its neighbor and rival China, might decide to pursue an arms buildup, too, matching the rate of the United States starting in 1954 (when the U.S. arsenal was similar in size to India's arsenal today). Pakistan, in turn (perhaps with support from China), might match India weapon-for-weapon. Again, these need not be strategically rational choices — political considerations and ideas about national glory can drive arms buildups. Suddenly, the world would find itself in a much worse position than even at the height of the Cold War:

arsenal on par with Washington's and Moscow's." Jakob Stokes, "Atomic Strait: How China's Nuclear Buildup Shapes Security Dynamics with Taiwan and the United States," *CNAS*, 1.

[187] **Detailed Explanation**: In this fictional scenario, China would reach 1500 warheads by 2035, as projected by the Department of Defense. In 2036, China then would have 1627 warheads, the same number that the USSR had in 1960. China's buildup from 2036 onward would then be an exact duplicate of the USSR's buildup from 1960 onward, culminating in 33,486 warheads in 2058 (the number that the USSR is estimated to have had in 1982). For each additional warhead added to China's arsenal in this scenario, one warhead is added to the arsenals of Russia and the United States.



Source: Author's visualization based on *Our World in Data* dataset and the scenario described above.

Again, this fictional scenario is not an implausible world. Arms-limitation treaties can be undone, as Russia's "suspension" of New START illustrates, and there are historical precedents for these kinds of arms races (after all, the weapon numbers in the graphs above are based directly on Cold War arming). The possibility of renewed arms racing should not be overstated, however. Some states — like North Korea — face severe resource constraints that differ from the constraints on the U.S. and Russia in the Cold War.[188] When states consider "how much is enough" for their nuclear deterrents, moreover, answers depend on objectives, which may not include matching their rivals weapon-for-weapon.[189] As described above, however, we are especially concerned about U.S.-China-Russia three-way arms racing dynamics, especially in light of statements by military planners.[190]

188 "[A]t least initially, new nuclear states usually face more acute resource constraints than the United States and the Soviet Union ever did while also facing normative, political, and legal constraints to their pursuit of a nuclear weapons capability." "How much is enough" in *The Fragile Balance of Terror,* 125.
189 "[In] defining enoughness, the place to begin is not capabilities but objectives. More narrowly, the notion of enoughness can be conceptualized as a threshold and is akin to, but not precisely, the notion of sufficiency." "How much is enough" in *The Fragile Balance of Terror*, 124.
190 See above, The Three Body Problem.

Questions for Further Investigation

- How valuable would more accurate information on possible nuclear winter effects be? How likely is it that additional funding would produce such information?
- Is it possible to produce better knowledge about civilizational collapse? What are the most important crucial considerations on these questions? How likely is such knowledge to be decision-relevant to funders and policymakers?
- How do different targeting policies — e.g. “no cities” — affect the cost functions described above?
- At what point do “overkill” and “rubble-bouncing” effects begin to level off the cost curve?
- At what point does nuclear winter-induced global cooling begin to outweigh the global warming caused by anthropogenic climate change?
- What are the biggest insecurities that could cause states like India to begin arms racing with China?
- How does adding or removing “rungs” from the escalation ladder shape the ultimate probability of reaching the most catastrophic rungs?

Guiding Principles for High-Impact Philanthropy

Key Points:

- Funders ought to focus on minimizing war damage. This ultimate goal *may* diverge from intermediate goals like disarmament, non-use, etc.
- In light of uncertainty about intervention effectiveness, funders ought to prioritize neglected strategies.
- Funders can multiply their impact by focusing on "great powers."
- Funders can multiply their impact by focusing on policy advocacy to leverage societal resources.
- The principle of "robust diversification" can help guide effective giving under the conditions described above.

The previous section established key facts about the structure of the nuclear risk problem that philanthropists are trying to manage. This section outlines guiding principles that can be derived from this problem structure and from an impact-maximizing approach to philanthropy.

Prioritize Minimizing Global War Damage

Founders Pledge's *Guide to the Changing Landscape of High-Impact Climate Philanthropy* relies on a fundamental insight: "the goal of high-impact climate philanthropy is not to maximize emissions reductions but to minimize climate damage."[191] A similar guiding principle applies to high-impact nuclear philanthropy; if we want to avoid human suffering and death, our goal should be to minimize expected global war damage. This goal is not necessarily identical to other goals often pursued by philanthropists and nonprofits, including:

- Disarmament and a nuclear weapons-free world;
- Avoiding the proliferation of nuclear weapons;
- Minimizing the probability of nuclear weapons use;
- Advocacy for arms control and against nuclear modernization.

[191] Johannes Ackva, Luisa Sandkühler, and Violet Buxton-Walsh, "A Guide to the Changing Landscape of High-Impact Climate Philanthropy" (Founders Pledge, November 2021), 12, 50, https://founderspledge.com/stories/changing-landscape.

To be clear, these objectives do not *necessarily* diverge from the goal of minimizing expected damage — better arms control, for example, may be desperately needed to avoid a three-way arms race between the U.S., Russia, and China — but they are not identical. They ought to be only viewed as *potential intermediate objectives* that may or may not serve the ultimate goal of minimizing expected global war damage.

If we combine prioritization of minimized expected global war damagewith the facts that war damages are non-linear and accidents are possible (as discussed above), we uncover one impact multiplier for high-impact philanthropy: *all else equal*, philanthropists ought to prioritize preventing larger nuclear wars over preventing smaller nuclear wars. This prioritization of larger wars may seem obvious, but it has major implications for marginal donations. It implies, for example, that work on escalation control — how to keep a limited war limited once it has begun — could be one of the highest-leverage interventions available. As discussed below (Philanthropy to the Right of Boom), it is also among the most neglected strategies.

Prioritize Neglected Strategies

We have already established a rough pattern in the importance of different kinds of nuclear wars, and deep uncertainty about the tractability of interventions. Given inefficiencies in the philanthropic market, *neglectedness* — how much attention and money an issue area receives — can now become a useful guide for understanding the relative impact of interventions or areas. As described below, political and ideological considerations may drive funder and grantee behavior, leading to undue neglect of different kinds of interventions. Categories of intervention that are neglected for the wrong reasons are likely to have more low hanging fruit. Thus, if philanthropists take a hits-based approach and are deeply uncertain about tractability, neglectedness becomes a critical impact multiplier. We discuss this further in the section on Philanthropy to the Right of Boom, below.

Is Everything Neglected Now?

Much of the analysis in this report is based in part on the insights from Founders Pledge's *Guide to the Changing Landscape of High-Impact Climate Philanthropy.* One key difference between the fields of climate and nuclear philanthropy, however, is their neglectedness; the billions spent annually on climate philanthropy dwarf the millions spent on nuclear security. In light of the funding shortfalls described above, we may ask, is *everything* neglected now? That is, is this report spilling unnecessary ink when what is required is a large-scale infusion of money into the sector?

In an ideal world, philanthropic attention to nuclear security (and to other global catastrophic risks) would increase by several orders of magnitude, but we clearly do not live in this ideal world. Therefore, it remains important to focus on *marginal* cost-effectiveness, especially for smaller donors. Moreover, insofar as there are misallocations in the overall philanthropic portfolio (discussed in the Right of Boom Philanthropy section below), now is an especially opportune time to correct them. Nonetheless, giving to general-purpose nuclear security organizations (such as the Ploughshares Fund or Nuclear Threat Initiative) may be a good option for some philanthropists who care about the health of the field, but disagree with the conclusions of this report.

Multiply Impact by Shaping Great Power Behavior

As explained in Founders Pledge's report on *Great Power Conflict*, there are strong reasons for impact-oriented philanthropists to focus on the behavior of the so-called "great powers."[192] This is for two reasons. First, great powers are simply the greatest potential originators of catastrophic nuclear risk — U.S.-Russia and future U.S.-China conflict would be catastrophic, and it therefore makes sense to prioritize these relations over the actions of minor nuclear powers like Israel. As discussed in *Great Power Conflict:*

> "Their military capacity allows Great Powers to compete with their rivals on the battlefield. It also allows them to affect the long-term future in a variety of ways: by facilitating cooperation or inflaming tensions, driving the development of destructive new technologies, or deploying highly-lethal weapons, including weapons of mass destruction."[193]

Second, great powers are able to lead and coerce weaker states to follow their lead. Fundamentally, this is about *leverage* over the world. Thus, for example, shaping the behavior of the United States is likely to affect the behavior of U.S. allies and partners, and alliances like NATO. The U.S. is therefore an especially large lever to an extent that smaller nuclear powers simply are not. These advantages are built into the rules of the international system, including through the UN Security Council and the veto power of the permanent members. These leverage factors make a focus on great power behavior a clear impact multiplier. (Because "great power" status is so fuzzy, this may be conceptualized as a simplistic binary multiplier.)

[192] Clare, *Great Power Conflict*.
[193] Ibid., 20.

Leverage Government Budgets via Policy Advocacy

As in climate philanthropy, nuclear philanthropists ought to target government action through *advocacy* for two reasons. First, nuclear policy — unlike many other targets of philanthropic interventions — simply is the purview of governments. Secrecy and classification make it impossible for most non-governmental organizations to have a *full* understanding of war planning, and governments ultimately make all the decisions about nuclear weapons.

Importantly, the massive defense budgets and resources of nuclear states provide another opportunity for leverage. Many billions of dollars are spent on the nuclear weapons complex by governments. To the extent that philanthropic action can affect the allocation of these billions, advocacy (broadly defined to include things like policy-relevant think tank work), can exert more leverage than even the largest philanthropic organizations could. The following case study provides an illustration of how this impact multiplier works.

Case Study on Leverage: Cooperative Threat Reduction

Cooperative Threat Reduction via the Nunn-Lugar Act provides one clear case study of leveraging government budgets via policy advocacy.[194] In the early 1990s, in the wake of the collapse of the Soviet Union, there was a concern that "loose nukes" in the former USSR would get into the wrong hands and lead to a massive problem of control and proliferation. Open Philanthropy (then known as GiveWell Labs), commissioned a report as part of its History of Philanthropy project that examined this case study.[195] As that report explains, catalytic grants by the Carnegie Corporation and the MacArthur Foundation ultimately helped lead to the elimination of over 7,500 nuclear warheads, plus additional thousands of delivery systems and nearly a thousand tons of chemical weapons.[196]

The specific mechanism of influence is relevant for understanding the power of leveraging government budgets. Carnegie and MacArthur funded Harvard Professor Ash Carter (later Secretary of Defense) to draft a report on *Soviet Nuclear Fission: Control of the Nuclear Arsenal in a Disintegrating Soviet Union*.[197] Senators Nunn and Lugar had already been

---

[194] For the purposes of this case study, the question of whether Nunn-Lugar reduced the risk of catastrophic nuclear war on net is irrelevant. Rather, it provides an example of a relevant *lever* that philanthropists can pull in order to multiply their impact.
[195] Benjamin Soskis, *Nunn Lugar Report* (2013), https://www.openphilanthropy.org/research/history-of-philanthropy-work-weve-commissioned.
[196] Ibid.
[197] Ashton B. Carter et al., *Soviet Nuclear Fission: Control of the Nuclear Arsenal in a Disintegrating Soviet Union* (Belfer Center for Science and International Affairs, 1991),

considering legislation on the problem of loose nukes, but the report — and Carter's scientific expertise and later advisory role in the drafting of legislation — likely catalyzed and accelerated the process at a moment when time was of the essence.[198]

We estimate that the philanthropic support behind this work was on the order of $5-10 million dollars over several years.[199] The governmental support of the 1991 Nunn-Lugar Act, in turn, was $500 million, and the overall amount spent by the U.S. government on cooperative threat reduction between 1992 and 2005 was over $5 billion.[200]

If we subjectively assign only 10% of the impact to Carnegie and MacArthur, and a $10 million philanthropic investment was 10% responsible for stimulating $5 billion in government funding, then there exists a 50-times multiplier of philanthropic money. Taking these numbers literally, this suggests ~75 nuclear weapons destroyed per million philanthropic dollars in expectation. Such large multipliers illustrate why the leverage of policy advocacy is crucial for reducing global catastrophic risks.

Robust Diversification

[This section is adapted from a previous Founders Pledge report on "Philanthropy to the Right of Boom."]

We have established two tentative points about nuclear philanthropy (which also hold for other international security questions):

1. Impact-oriented philanthropists ought to prioritize preventing the largest wars, all else equal;

---

https://www.belfercenter.org/publication/soviet-nuclear-fission-control-nuclear-arsenal-disintegrating-soviet-union.

[198] Paul Rubinson summarizes this point from Soskis: " Although Soskis notes that Senator Nunn may have pursued such a program without the influence of philanthropy, he argues that philanthropy at least catalyzed the process. This seems the likeliest scenario, as it is reflected in other literature about cooperative threat reduction. After all, it took very little to convince politicians that loose nukes were a threat to US national security—the dilemma was what to do about it. As Soskis recognizes, giving money directly to Russia looked bad, but the data and recommendations crafted with philanthropic grants were so specific and convincing (perhaps even to the point of being incorporated into the wording of the legislation) as to demonstrate that money going to Russia would serve US national interests." Paul Rubins on, "Philanthropy, Nuclear Nonproliferation, and Threat Reduction," *Urban Institute* (2021), 8.

[199] As Soskis notes, the exact amount is difficult to establish, given the collaboration of two funders and the diffuse network of experts and think tanks involved, but in 1989, MacArthur made a $5 million grant to Brookings for a 5 year project, which involved collaboration with the Center for Science and International Affairs at Harvard, Stanford, and Carnegie Endowment (Benjamin Soskis, *Nunn Lugar Report* (2013), https://www.openphilanthropy.org/research/history-of-philanthropy-work-weve-commissioned.) If we assume that Carnegie support was on a similar order of magnitude, the estimate of $5-10 million appears plausible.

[200]Ibid and https://www.govinfo.gov/content/pkg/GAOREPORTS-GAO-05-329/html/GAOREPORTS-GAO-05-329.htm

2. There is high uncertainty on the effectiveness of specific interventions, and often no way to rigorously compare them.

As mentioned above, we can thus look to *neglectedness* as a potential impact multiplier and pursue a strategy of "robust diversification" to better prioritize interventions within the nuclear field. We are borrowing this concept of robust diversification from Founders Pledge's *Guide to the Changing Landscape of High-Impact Climate Philanthropy*.[201]

Robust diversification has two components:

1. "**Portfolio diversification with negative correlations**" — "When deeply uncertain about the precise returns of different strategies, we combine strategies where the uncertainties are negatively correlated, so that when one uncertainty is resolved 'pessimistically' chances are the other uncertainties are resolved positively."[202]

2. "**Robustness to the worst worlds"** — Given assumptions about non-linear damages (which apply for nuclear war, where climate effects could make the largest wars disproportionately worse than smaller wars), we ought to "prioritize strategies that are effective under pessimistic assumptions."[203]

In practice, robustness to the worst worlds means understanding how to reduce nuclear risk "when deterrence fails" — that is, when a nuclear war breaks out. Similarly, in a great power conflict, we focus on all-out "total war" with massive targeting of cities and civilians (especially if power-law distribution assumptions about wars hold[204]). As the next section suggests, philanthropists focusing on nuclear issues appear to disproportionately neglect these very topics.

---

[201] Johannes Ackva, Luisa Sandkühler, and Violet Buxton-Walsh, *A guide to the changing landscape of high-impact climate philanthropy,* Founders Pledge, 2021, https://assets.ctfassets.net/x5sq5djrgbwu/7eEpX4UcKNEy6LUDhf2B05/735518c277987ad5ad91f096b1fdc2a7/A_guide_to_the_changing_landscape_of_high-impact_climate_philanthropy.pdf.
[202] Ackva et al., *Guide to the Changing Landscape of High-Impact Climate Philanthropy,* 66.
[203] Ibid., 67.
[204] "Power-law distributions are the key characteristics of the highly improbable, unimaginably large phenomena that Nassim Taleb described in his bestselling book *The Black Swan.* If an outcome follows a power-law distribution, the overwhelming majority of the observations of that outcome are small, but a few are really, really huge." (Bear F. Braumoeller, *Only the Dead: The Persistence of War in the Modern Age,* 104.) For a discussion of the severity of war as a power law distribution, see *Only the Dead* and Lars-Erik Cederman, T. Camber Warren, and Didier Sornette, "Testing Clausewitz: Nationalism, mass mobilization, and the severity of war," *International Organization* 65, no. 4 (2011): 605-638.

Questions for Further Investigation

- How many traditional philanthropic actors (if any) explicitly share the goal of minimizing the expected damage of nuclear war?
- Can philanthropists quantify the benefits of prioritizing great power behavior, or should this be treated as a binary multiplier?
- What organizations have the most leverage over government policies on specific areas of nuclear risk?
- How should philanthropists balance the benefits of “catalytic” grants to grow small organizations with the benefits of greater policy influence by more established organizations?
- What other issue areas would benefit from “robust diversification”?

Philanthropy to the Right of Boom

[This section is adapted from a previous Founders Pledge report on "Philanthropy to the Right of Boom." Readers familiar with that document may wish to skip this section.]

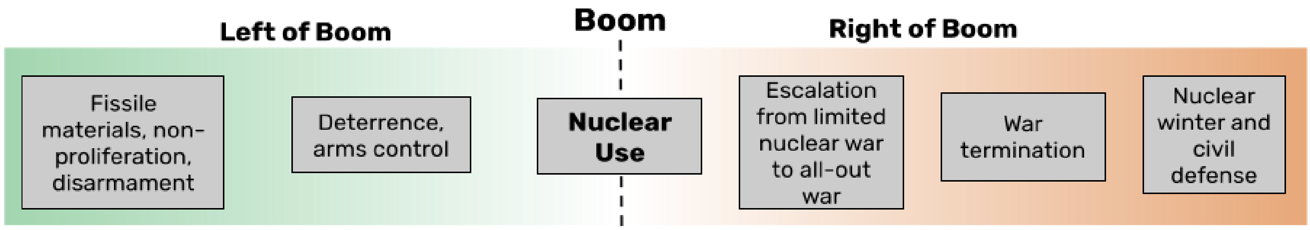
Key Points:

- "Right of boom" interventions — focusing on problems arising *after* the first use of nuclear weapons, such as escalation control — are an important part of risk reduction;
- These very interventions have been severely neglected by philanthropic funders, possibly for ideological reasons;
- These facts suggest that prioritizing "right of boom" interventions is a promising impact multiplier for funders.

Right-of-Boom Neglectedness

The field of nuclear risk reduction can be overwhelming, and the option space of possible interventions is large. The considerations outlined in the previous sections, however, point to an important distinction between interventions that helps to categorize possible funding opportunities: whether an intervention acts to the "left" or "right" of "boom" — before or after nuclear first use.

The diagram below roughly outlines the difference between "left" and "right" of boom interventions. In short, interventions that seek to act *before* the first use of a nuclear weapon are "left" on the spectrum (akin to "left of launch"), and interventions that seek to act *after* the first use of a nuclear weapon are "right." The neatness of this division is artificial, and in practice the differences are fuzzy, but it points towards a real distinction.



Left of Boom	Right of Boom
• Deterrence • Conventional escalation management and war limitation • Conventional war termination • Non-proliferation • Disarmament • Some arms control • Peace activism/advocacy • Non-use norms/"taboo" • Efforts to reduce probability of nuclear use in conventional conflict	• Deterrence (e.g. of further escalation) • Nuclear escalation management and nuclear war limitation • Nuclear war termination • Deterrence failure and fighting nuclear war • Civil Defense (including food stockpiling and alternative food production) • Hotlines (war limitation or termination)[205] • Understanding and mitigating climate effects of nuclear war ("nuclear winter") • Planning for post-war political environment • Missile defense

Source: Author's visualization.

[205] See Founders Pledge's recent report on hotlines and crisis communication systems, *Call Me, Maybe? Hotlines and Global Catastrophic Risks*, https://docs.google.com/document/d/1U2sVOLYhf7uejXd08HDylu3p0smd0ES7ru0xhcUaoJI/edit.

To re-emphasize, not all nuclear war is equally damaging and, for impact-minded philanthropists, almost all of the cost of nuclear war lies well to the right of boom on the spectrum presented here.[206] This is especially true if, as seems likely, several of the main plausible existential risks — civilizational collapse and extreme nuclear winter — only obtain at the far right end of the spectrum. The key question therefore becomes: after nuclear first use, how can nuclear war and its damage be limited?

Funding right-of-boom interventions does not necessarily mean only funding food stockpiling or bunker-building. A "right-of-boom" intervention could be a large-scale research project on the problem of how to keep limited nuclear war from escalating into all-out war. As explained above, this distinction — between small and large wars — is likely among the most salient distinctions in nuclear issues for anyone interested in maximizing impact. Such a research project may find that limited war is unlikely to remain limited (e.g. there are historical base rates on escalation in other contexts, or in well-designed wargames), and thus make other right-of-boom interventions appear less promising, or it may open up new lines of research into escalation management.

Moreover, some interventions work on both sides of "boom."[207] For example, deterrence is important both left and right of boom, e.g. to deter further escalation from a limited nuclear war. Similarly, war termination (of a conventional war) can be an important left-of-boom intervention — this may include better abilities to negotiate the terms of a peace treaty. Interventions on one side can strengthen those on the other. In short, the distinction drawn here can be useful, but we should be careful not to oversimplify the categories.

Finally, right-of-boom interventions need not be unilateral. As was recognized by some in the Cold War, right-of-boom work like civil defense can in fact be conceptualized as a kind of risk-reduction exercise analogous to arms control, which can be pursued in bilateral or even multilateral fora, track II diplomatic discussions, and more. For example, states can have technical exchanges on how to best protect civilian populations in the event of a nuclear war.[208]

---

[206] Some people may view the very existence of nuclear weapons as immoral (e.g. because their existence perpetuates power imbalances), may see any "nuclear war" as fundamentally wrong, or may not distinguish degrees of badness between different kinds of nuclear wars. This report prioritizes minimizing expected cost in terms of death and suffering. .
[207] Thanks to James Acton for pointing to this, and for each of the points in this paragraph.
[208] "In the United States, in the mid-1960s assistant secretary of defense for civil defense Steuart L. Pittman argued that civil defense should be considered a form of arms control. He suggested that cooperation with the Soviet Union on civil defense, beginning with technical exchanges and possibly building up to steps such

The following sections ask three questions. First, is the apparent neglect of "right of boom" interventions a real phenomenon? Second, if it is real, is this neglect based on evidence or otherwise rigorously justified, or does it reflect other biases (e.g. political ideology)? Third, how should new funders interested in the nuclear field approach the problem of "right of boom philanthropy"?

Quantifying Right-of-Boom Philanthropy

To begin to answer these questions, we searched through all the grants in the subject area "Nuclear Issues" of the Peace and Security Funding Index and identified any grants that may be considered "right of boom" based on the descriptions in the database.[209] For ease of reference, we have included the relevant grants in the appendix.[210] To be conservative, we also counted those grants where it was unclear whether they ought to be considered right of boom. The analysis suggests that the phenomenon identified anecdotally is real: **right-of-boom projects receive at most one-thirtieth of the total funding of the nuclear field.**[211]

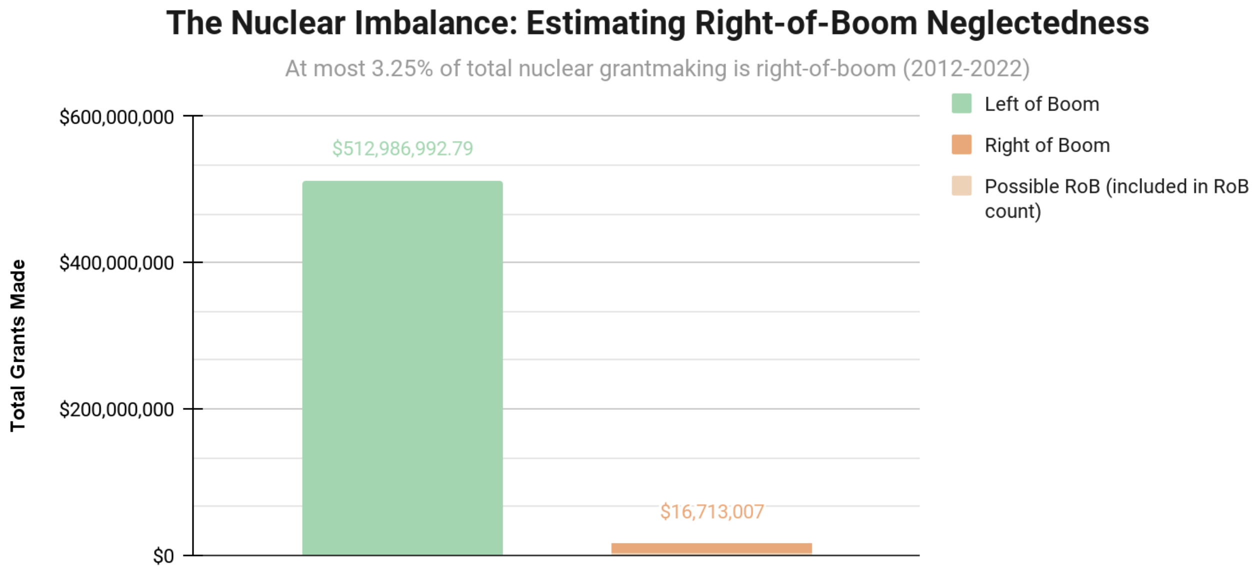
---

as the mutual elimination of megaton-range weapons on humanitarian grounds, could not just make war less destructive if it occurred, but make it less likely in the first place." Edward M. Geist, *Armageddon Insurance: Civil Defense in the United States and Soviet Union, 1945-1991*, 7.

[209] Candid, *Peace and Security Funding Index*, accessed 25 October 2022. N.B., this may fail to capture some grants (especially outside the United States), but we believe it gives a relatively accurate overview of the field. For more information on how the data were collected, see Candid, "Data Sources," candid.org, accessed 25 October 2022.

[210] Additionally, we cross-referenced this with a search of Open Philanthropy's grants database and information found on other funders' websites, as well as other projects, such as ALLFED, that we know are working on right-of-boom interventions.

[211] This process involved reading through many grant descriptions, and it is possible that the analysis is mis-labeling some grants and missing others, including those that are not reported in the Peace and Security Funding Index. This is intended as a first attempt at estimating the magnitude of the imbalance.



Source: Author's estimate, using data from *Peace and Security Funding Index* and other sources.

Because of the inclusiveness of the search, 1-in-30 is likely an *upper* bound estimate (i.e., the actual share is likely much smaller).[212] Nuclear subject-matter experts interviewed for this project also suggested that 1-in-30 intuitively seems too high. For example, James Acton, the co-director of the Carnegie Nuclear Policy Program, said "1-in-30 is an upper bound [...] I have to say I would be pretty surprised if it was as much as 1-in-30."[213]

Government Work Right of Boom

Complicating this picture, some government agencies do consider right-of-boom interventions. For example, U.S. Strategic Command (STRATCOM) is tasked in part with understanding how a nuclear war would actually be fought, and some parts of STRATCOM may spend substantial time on right-of-boom interventions.[214] Public traces of STRATCOM's war planning analyses and their right-of-boom focus can be found in e.g. the *Report on the Nuclear Employment Strategy*:

> "to strengthen the credibility of U.S. nuclear deterrence and extended deterrence, the United States will continue to field a range of nuclear and nonnuclear capabilities

---

[212] Several developments in nuclear philanthropy make this a generally unstable estimate. For example, the Future of Life Institute has put out a [request for proposals] to study the consequences of nuclear war — a right-of-boom question — and the MacArthur Foundation has shrunk its nuclear grantmaking and will make its [final grants] in the field in 2023.
[213] Call with James Acton and Ankit Panda, 10 January 2023.
[214] Thanks to James Acton for this point.

that provide U.S. leadership with options that can be tailored to deter potential adversaries, assure allies and partners, achieve U.S. objectives should deterrence fail, and hedge against an uncertain future."[215]

Specifically, phrases such as "If deterrence fails, the United States will strive to end any conflict at the lowest level of damage possible and on the best achievable terms for the United States, and its allies, and partners" and "elements of U.S. nuclear forces are intended to provide limited, flexible, and graduated response options" hint at the importance of right-of-boom planning in U.S. nuclear employment strategy.[216]

I have so far been unable to quantify the amount of within-government right-of-boom work, and this may be impossible, given classification issues. Nonetheless, government work in some parts of the U.S. government — parts of STRATCOM, parts of Homeland Security, etc. — leans more heavily to the right of boom than some non-governmental organizations do.[217] It is tempting to suggest that non-governmental spending left of boom is actually a corrective to governmental bias for right-of-boom interventions. In other words, the private sector is providing more balance to the overall analysis of nuclear issues. As discussed below, however, I do not think that there really is a governmental bias for right-of-boom interventions.[218] Rather, there appear to be merely pockets of the government that truly see right-of-boom analysis as part of their portfolio. Government neglect of such critical issues related to nuclear war, though baffling, is not historically unusual. One anecdote about Herman Kahn (via Paul Bracken) illustrates this:[219]

> "In a Pentagon briefing in the 1970s, Herman Kahn audaciously proclaimed that he and his colleagues at the Hudson Institute happened to be the world experts on ending a nuclear war. The Pentagon had studied many ways that a nuclear war could start, Kahn argued, but not how it would end. The audience was incredulous. One official challenged Kahn as to how anyone could possibly be an expert on ending a nuclear war. Kahn shot back: 'I put two junior people on it for a couple days last week. We've thought more about it than the entire Department of Defense has.'"[220]

---

[215] *Report on the Nuclear Employment Strategy of the United States – 2020,* https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/NCB/21-F-0591_2020_Report_of_the_Nuclear_Employement_Strategy_of_the_United_States.pdf.
[216] Ibid.
[217] Though STRATCOM's 2023 research topics for its Academic Alliance do not seem to reflect this.
[218] This analysis is notably limited to the United States, as a major nuclear power that is relatively transparent about nuclear weapons and strategy.
[219] Thanks to Matthew Gentzel for pointing me to this quote.
[220] Paul Bracken, "Net Assessment: A Practical Guide," *The US Army War College Quarterly: Parameters* 36, no. 1 (March 1, 2006): 95, https://doi.org/10.55540/0031-1723.2285.

As Bracken writes, “To get an estimate wrong, there has to be an estimate in the first place. What one often finds, unfortunately, isn’t estimates that are wrong, but the realization that no one has thought about any estimates at all.”[221] On many issues right-of-boom, it appears that either no one has thought about them at all, or no-one has thought about them since the early days of the Cold War.

The limited governmental interest that does exist for these topics may, moreover, give non-governmental right-of-boom analyses and interventions more leverage over government thinking and allocations.[222] Non-governmental analyses may also provide useful correctives to biased thinking in government. Secrecy, organizational culture, self-selection, and other sources of bias may fundamentally distort the government’s analysis of right-of-boom interventions, and non-governmental analysis can provide new ideas, balance, criticism, and corrections. For example, non-governmental analysis might focus on the wellbeing of humanity rather than the U.S. national interest (to the extent that these may diverge). Moreover, STRATCOM’s war planners likely take a specific “warfighting” lens to right-of-boom planning that may not explore the full spectrum of right-of-boom interventions.[223]

Finally, government interest in right-of-boom work may increase the policy leverage of non-governmental right-of-boom analyses. James Acton explained this perspective: “The fact that STRATCOM is working on this, for me, is an argument for why we *should* work on it. Because it's an issue that the government cares about, and it's a lot harder to make policy change when the government is not thinking about an issue than when it is.”[224] In other words, just as climate philanthropy ought to leverage government action to have the highest possible impact, so nuclear philanthropy can leverage government interest in right-of-boom interventions to maximize its impact.[225]

---

[221] Ibid.
[222] Thanks to James Acton for pointing to this.
[223] “STRATCOM approaches [...] Right of Boom through a very particular lens. It's a warfighting lens. It's ‘how do we win the war?’ And, you know, as military officers, you think about things in terms of winning and losing wars, essentially, by overpowering your adversary. You're not necessarily trained in escalation and escalation management and the psychological and perceptual dimensions of all of this and the compellent/coercive dimension.” (James Acton, call with James Acton and Ankit Panda, 10 January 2023)
[224] Call with James Acton and Ankit Panda, 10 January 2023. This depends on which parts of the government one is considering, and specifically, how politically salient an issue is.
[225] “We firmly believe that — on balance —funding advocacy efforts to induce policy change and affect how societal resources are spent, provides the most compelling proposition for impact-oriented philanthropists.” (Ackva et al., *Guide to the Changing Landscape*).

Possible Explanations for Neglectedness

Neglectedness is not by itself a reason to fund a class of interventions. Rather, philanthropists ought to care about *undue* neglectedness — is an issue or intervention neglected relative to its importance and tractability, and is it being neglected for the wrong reasons (i.e. reasons that do not affect the cost-effectiveness of donations)?[226] Therefore, this section explores possible explanations for why traditional philanthropists do not focus on right-of-boom nuclear interventions.

Explanation 1: Differing Worldviews and Moral Priorities

It would be uncharitable to say that most funders and grantees in nuclear weapons-related philanthropy are "wrong" to neglect right-of-boom interventions. Rather, the prioritization framework that would lead one to focus on the most extreme kinds of nuclear wars is simply unintuitive to many people. The idea that some nuclear wars are worse than others, for example, may not make sense to people with non-consequentialist views. Funders and grantees might believe, e.g.:

- The very existence of nuclear weapons is immoral, even if they are never used;[227]
- Disarmament and non-proliferation are ends in themselves (rather than means to prevent suffering);
- The point of nuclear weapons philanthropy and advocacy is to correct past and present injustices.[228]

Holding any of these views might lead funders to endorse the kind of allocation described above.

Explanation 2: Right-of-Boom interventions are Politically/Ideologically Unpalatable

Relatedly, there are political or ideological reasons why funders may neglect right-of-boom interventions. Few non-governmental analysts and scholars are incentivized to conduct right-of-boom research. James Acton explained that, "It's an issue where for very different reasons, ideologically few people have an interest in doing this work" and that "people on the [political] Left don't want to acknowledge there is a possibility of controlling nuclear war, because that could be an argument for the nuclear-armed SLCM [submarine-launched

---

[226] In the words of one description of this framework, "neglectedness is only a good proxy if the area is being neglected for bad reasons by other actors." ("Problem Framework, *80,000 Hours,* https://80000hours.org/articles/problem-framework/#how-to-assess-how-neglected-a-problem-is)
[227] This seems to be the view of the Catholic church and of many disarmament activists (CNA, "Pope Francis: Nuclear Weapons Are 'Immoral,'" Catholic News Agency, https://www.catholicnewsagency.com/news/251596/pope-francis-nuclear-weapons-are-immoral).
[228] See, e.g. "Racism and Nuclear Weapons," ICAN, accessed December 8, 2022, https://www.icanw.org/racism_and_nuclear_weapons.

cruise missile] or the low-yield D5 [a modified warhead for a Trident submarine-launched ballistic missile] [...] and people on the Right don't want to do this, because they look like Dr. Strangelove, and everyone thinks they're crazy."[229]

This may stem from the historically hawkish political associations of Cold War right-of-boom research, which to some appeared to downplay nuclear risk and accelerate the arms race.[230] More broadly, the very idea of escalation management may simply have appeared outdated in the post-Cold War environment. As Ankit Panda explained, "For much of the last 30 years post-Cold War, the idea of studying escalation management in a nuclear war was just not where the times were taking us. Nuclear arsenals were declining. The idea of nuclear war was broadly kind of pushed aside."[231] Yet even in the Cold War, much right-of-boom thinking was neglected relative to the resources invested in the nuclear weapons complex and the risks taken by nuclear states. As Edward Geist, a historian and policy researcher at the RAND Corporation, has written about the relative neglect of civil defense, "both the Soviet and American governments were willing to risk the destruction of civilization, yet saw comparatively little reason to try to save it if they came to blows."[232]

Explanation 3: Right-of-Boom Interventions Present PR/optics Problems for Funders.

Related to the points above, these kinds of interventions are also difficult for donors to justify to the public, especially if they want to appear to be focused on peace. A foundation's board, in particular, may object to funding right-of-boom interventions on grounds of the historical associations of such work with hawkish Cold Warriors.[233]

Explanation 4: Right-of-Boom Interventions are too Technical for some Funders to Understand

These kinds of interventions are often more technical than other nuclear interventions. Funding grassroots campaigns to protest nuclear weapons is easy to understand; making

---

[229] Call with James Acton and Ankit Panda, 10 January 2023.
[230] As arms control expert Jeffrey Lewis explained when reviewing an earlier draft of this document, "This work has, in the past, been tainted by its popularity with a certain group of people. [Herman] Kahn's emphasis on escalation control and enthusiasm for civil defense were associated with a general framework that appeared to drive the arms race and neglect interventions that reduce risk. That is not to say that such work could not be done with real integrity." (Google Docs comment by Jeffrey Lewis, 13 January 2023)
[231] Call with James Acton and Ankit Panda, 10 January 2023. Similarly, Acton: "It's a problem that's attracted so little attention, especially since the end of the Cold War. It seems to me that there's real potential upside to spending time and effort thinking about those problems."
[232] Geist, *Armageddon Insurance,* 8.
[233] Thanks to Jeffrey Lewis for pointing out that boards may present a bigger obstacle for many funders than public perception *per se.*

sure limited nuclear use does not escalate to all-out war may involve game theory or details about the yields of various weapons.

Explanation 5: Funders Think Right-of-Boom Interventions are Intractable

Another possible explanation is that right-of-boom interventions are fundamentally intractable in some sense. For example, a funder might believe that nuclear war, once begun, is *almost certain* to escalate to all-out war, such that an attempt to limit nuclear war would be futile. Similarly, attempts at civil defense have long been ridiculed as futile in the face of the overwhelming threat of nuclear war.[234] There are two problems with this reasoning, however.

First, we simply don't know whether this is true. It could be the case that war limitation is highly tractable. There has been no nuclear war, and only highly theoretical work on escalation, so we do not know how intractable these interventions really are.

Second, we want to be prepared for the possibility that left-of-boom interventions fail; this is part of the reasoning behind robust diversification. The history of "near misses" in the Cold War suggests that accidents simply do happen. The world ought to be prepared for the scenarios where we find ourselves in a nuclear war, and want to know how to limit the damage. Accidental nuclear war has almost happened several times in history; we need to make sure that every accident does not lead to the destruction of modern civilization.

Explanation 6: Funders Think Right-of-Boom Interventions are Dangerous

A second explanation — related to Explanation 2 — is that right-of-boom interventions are in some way dangerous. For example, one argument against studying how to keep nuclear war limited is that doing so would itself make nuclear war seem "winnable" and thus weaken the nuclear taboo. Similarly, working on civil defense and resilience interventions, the argument goes, would make the war seem "survivable" and thus make war more likely. Again, this fails to account for the fact that accidents and unintended escalation do happen, and need to be insured against; rationality is not the only thing governing whether or not the world goes to war. Second, this argument neglects the possibility that e.g. civil defense interventions increase the attacker's uncertainty about the effects of their weapons, thus potentially making nuclear use *less* likely.

[234] As mentioned in the introduction, one popular joke in Cold War America asked, "What do you do when you see a flash? You put your head between your legs and kiss your ass goodbye." (quoted in Geist, *Armageddon Insurance,* 11)

The overall concern about moral hazard and informational hazards is valid. It is also the very concern that would be explored by questions that some people do not wish to ask, and that more funding on escalation management could help to answer.

Explanation 7: Grantees Are Not Interested in this Work

The previous six explanations have focused on the perspective of funders, but we should note that there may also be non-funder-related reasons for why right-of-boom projects do not get funded — grantees may simply be uninterested.[235] Analysts at think tanks and scholars at universities may share several of the views outlined above, and may therefore not submit right-of-boom grant proposals in the first place. Moreover, a dynamic may ensue where analysts assume funders are uninterested in this work, and funders in turn are unable to find worthwhile grantees, thus making fewer grants and reinforcing the idea that traditional philanthropists won't fund this kind of project. Once again, however, this does not necessarily reflect rational reasons for the neglect of right-of-boom interventions.

Explanation 8: Funders and Think Tanks Take Their Lead from Government Interests[236]

Finally, the non-governmental nuclear space interacts with the government in many ways, including conferences, other think tank events, and the movement of people into and out of government. The predominance of left-of-boom work in much of the nuclear enterprise (such as in the Departments of State and Energy, the National Security Council, Congress, and multilateral organizations like the IAEA) stands in contrast to the relatively few organizations who see right-of-boom work as explicitly part of their portfolio (STRATCOM, Homeland Security, and parts of National Laboratories, among others). The influence of this apparent imbalance on foundation staff and boards may in turn shape non-governmental work.

Robust Diversification with Right-of-Boom Interventions

This write-up outlines several tentative observations about the landscape of nuclear philanthropy:

1. Traditional philanthropy appears to neglect right-of-boom interventions;
2. This is not necessarily for rational reasons, but may reflect funder biases or non-consequentialist worldviews;

[235] Thanks to James Acton for this point.
[236] Thanks to Carl Robichaud for this point.

3. Right-of-boom interventions are the very interventions that might keep a small nuclear war from becoming a civilizational-collapse event.

The 30-to-1 neglectedness of these kinds of interventions, combined with these facts, seems to suggest that we ought to consider the right-of-boom distinction as a potentially promising impact multiplier.

This analysis suggests that attention and political preferences skew interventions in such a way that makes robust diversification the best strategy for effective philanthropy *on the margin* for nuclear philanthropy, as it does for climate philanthropy. We expect similar kinds of biases to be found in other cause areas, but we have not investigated this further. On great power war more broadly, the difference between border skirmishes and firebombing cities is similarly important — how can we prevent the worst outcomes and decrease the likelihood of escalation?

This analysis is intended as a first approach to the problem of right-of-boom philanthropy. Additional research on these kinds of interventions may be especially important in light of U.S.-China competition and conflict. Whereas research on limited war between the U.S. and Russia goes back to the early years of the Cold War, scholars of Chinese nuclear strategy have suggested that Chinese experts hold beliefs about the (un)controllability of nuclear escalation that differ fundamentally from the beliefs of U.S. experts and military strategists.[237] These differences may be a major factor shaping the likelihood of nuclear war and escalation dynamics within such a war.[238]

Despite the preliminary nature of this analysis, however, the possibility of a 1-to-30 (and likely much higher[239]) impact multiplier helps to narrow the field of possible high-impact interventions. All else equal, philanthropists looking to maximize the relative effectiveness

---

[237] "…if nuclear weapons are used in a conflict—however limited that initial use—they [Chinese experts] believe that subsequent escalation would not be controlled, which restrains leaders from pursuing even limited use. […] China's skepticism about controlling nuclear escalation is reflected in its nuclear doctrine, and force structure, which is not optimized for conducting limited strikes." Fiona Cunningham and M. Taylor Fravel, "Dangerous Confidence? Chinese Views on Nuclear Escalation," *International Security* 44, no. 2 (2019): 64.
[238] "[I]n a crisis or conflict with China, the United States might overestimate the likelihood that China would use nuclear weapons and underestimate the scale of a Chinese retaliatory nuclear strike. Paradoxically, then, Chinese views about nuclear escalation may be suboptimal from the perspective of China's ability to deter either nuclear attacks or conventional attacks on its nuclear arsenal by the United States and create greater instability during a crisis." Ibid.
[239] For comparing the relative expected impact of different interventions, this may be best represented as a distribution, rather than a point estimate.

of their donations in reducing nuclear risk ought to leverage this neglectedness multiplier when choosing where to give.

Questions for Further Investigation

- Are right-of-boom interventions similarly neglected in other countries?
- How much more neglected than 30-to-1 are right-of-boom interventions really?
- How does this ratio compare to the focus in government organizations like STRATCOM?
    - Can proxy indicators (e.g. Minerva Grants) help us establish this ratio using publicly-available sources?
- Should philanthropists factor PR concerns into their impact calculations? If so, how?

Multiplying Impact and Sample Interventions

Key Points:

- Funders can use actionable "impact multipliers" from the previous sections to guide their giving;
- The features described in this report point towards concrete interventions that philanthropists can fund.

This section translates the theoretical discussions of the earlier parts of this report into actionable interventions that philanthropists could fund. This is not intended to be an exhaustive list, but to show how philanthropists can use the impact multiplier framework to guide effective giving. As a reminder, key ways to multiply impact include:

- **Prioritize larger nuclear wars** and focus on minimizing damage;
- **Focus on neglected strategies**, especially "right-of-boom" interventions;
- **Target "Great Power" behavior and policy**;
- **Leverage government budgets via advocacy**;
- **Pursue a strategy of "robust diversification"** that hedges against worlds where standard approaches have failed.

Large-Scale Research and Policy Project on the "Three Body Problem" of Nuclear Deterrence

Our highest-priority recommendation for effective philanthropy is to fund a large-scale multi-year research project on the "nuclear three body problem." Such a research project would seek to study key questions of three-way deterrence and arms control:

- How must U.S. nuclear policy change to maximize strategic stability when competing with two major nuclear powers?
- If deterrence fails and nuclear war breaks out between any of the major nuclear powers, how can war be limited and controlled to minimize damage?
    - Note that this question often cuts directly against the first.
- What can game-theoretic models tell us about this problem? What other methodological approaches can help to illuminate the problem of managing this relationship?

- What, if any, Cold War bipolar frameworks are still applicable in a three-body world? How can others be redesigned?
- What does three-way deterrence look like in 2027, 2030, 2035? What are the expected capabilities? How can they be forecast more accurately?

Based on interactions with experts and potential grantees and previous grants for policy-relevant research, we estimate that a high-quality smaller project could be funded at a single institution for about $250,000 to $500,000 per year, but that an investment on the order of $1 million to $10 million per year would have a more transformative impact. This could be spread among think tanks and academic institutions. Although most think tanks bill themselves as "nonpartisan," this group ought to include traditionally hawkish and conservative institutions, in order to increase the chance of influencing future Republican administrations. Given our uncertainty about these issues, an approach of seeding multiple papers and studies at various institutions is more appropriate than funding a single institution at massive scale.[240]

In addition to research, such a project could include track II and track 1.5 (unofficial with some government participation) dialogues between the United States, Russia, and China. As far as we are aware, there is currently only one track II dialogue (hosted by the Pacific Forum) that is explicitly focused on U.S.-China strategic nuclear issues (although many others touch on this topic intermittently).

If the research can be coupled with high-quality policy advocacy (closed-door policy briefings, op-eds in major newspapers, Congressional testimony), then funding of such a project could **hit all the relevant impact multipliers identified above**.

Reinvigorating "Right of Boom" Thinking

A second group of potential high-impact interventions for philanthropists to fund would be various policy-relevant "right of boom" analyses. This could be structured as a Request for Proposals, to cast a wide net for potential projects. Projects could be selected based on how well they hit the impact multipliers identified in this report.

One especially promising subset of "right-of-boom" work is the idea of "winter-safe deterrence," broadly defined.[241] By this, we do not necessarily mean extreme arsenal reductions or the substitution of other strategic weapons, both of which have been

---

[240] Thanks to James Scouras for this point in a round of external review.

[241] The term "winter-safe deterrence" was coined by Seth Baum, but is used here without reference to other types of weapons (Seth D. Baum, "Winter-Safe Deterrence: The Risk of Nuclear Winter and Its Challenge to Deterrence," *Contemporary Security Policy* 36, no. 1 (January 2, 2015): 123–48, https://doi.org/10.1080/13523260.2015.1012346.)

suggested as "winter-safe" approaches.[242] Rather, we want to know whether there exist nuclear postures and capabilities that retain potential deterrent effects while minimizing the probability of nuclear winter. For example, can yields, burst heights, targeting policies, etc. be adjusted in such ways as to minimize soot production while retaining deterrence?[243] Additional considerations may include "civil defense" broadly defined, as well as the problem of recuperation after nuclear war.[244]

Foundational Studies and Challenging Conventional Wisdom on Nuclear War[245]

A third avenue for impact is to fund several policy-relevant "foundational studies" and to challenge the "conventional wisdom" on nuclear war — with special emphasis on projects that seek to identify crucial considerations on great power behavior, minimizing expected war damage, and rigorously prioritizing possible policy options. As illustrated in this report and elsewhere, some commonly-held assumptions about nuclear strategy are simply false. At times, citations lead to more citations that eventually lead to speculative analyses from the 1950s, often justified with pseudo-scientific early Cold War "facts" about human psychology or decision-making. Recent analysis by the Johns Hopkins Applied Physics Laboratory, for example, critically examined a family of charts claiming to illustrate "wartime fatalities in the nuclear era" — charts widely used by senior U.S. defense leaders and in critical policy documents — and found that the so-called graphs are irreproducible with any real data, statistically misleading, and ultimately traceable to nothing but cartoonish drawings.[246] The field not only holds on doggedly to some dubious theories, but generally lacks diversity and new thinking; an influx of epistemic diversity may be beneficial. Foundational studies may include:

- Understanding the implications of uncertainty around nuclear winter for policy options;[247]
- Detailed analyses of neglected post-Cold War nuclear crises and how they differ (or not) from better-understood Cold War case studies;

[242] Ibid.
[243] Thanks to Dr. James Scouras for this point.
[244] Thanks to Dr. James Scouras for this point.
[245] This entire section is indebted to detailed comments by Dr. James Scouras in a round of external review.
[246] Lauren Ice, James Scouras, and Edward Toton, "Wartime Fatalities in the Nuclear Era (National Security Report)," Johns Hopkins Applied Physics Laboratory, https://www.jhuapl.edu/sites/default/files/2022-12/WartimeFatalities.pdf.
[247] For an example of this kind of work, see James Scouras, Lauren Ice, and Megan Proper, "Nuclear Winter, Nuclear Strategy, Nuclear Risk (National Security Perspective)," Johns Hopkins Applied Physics Laboratory, https://www.jhuapl.edu/sites/default/files/2023-05/NuclearWinter-Strategy-Risk-WEB.pdf.

- Developing new rigorous analytic frameworks for decision-making under extreme uncertainty — possibly including the integration of probabilistic forecasting — for understanding whether a given policy option is likely to do more harm or good.

Generally, any funding program supporting such studies could follow the model outlined in the previous two sections, and could be designed to specifically prioritize the impact multipliers identified in this report.

Other Interventions

- **Chinese views on escalation control;**
    - Chinese views on nuclear escalation remain poorly-studied, with only a small number of academic articles published on the subject.[248]
- **Escalation Control on the Korean Peninsula;**
    - How can nuclear war on the Korean peninsula be limited in intensity and scope, if it were to break out?
    - What agreements can the United States and China reach around these risks?
- **Escalation Control in South Asia;**
    - What are the risks of unintended escalation and nuclear use in South Asia?
    - What are the risks that India-Pakistan war would draw in outside powers, including China and the United States?
- **Shaping Indian nuclear policy;**
    - What are the biggest holes in Indian civil society related to nuclear policy?
    - What are the most effective ways to shape India's rise in a way beneficial for international stability?

[248] Fiona S. Cunningham and M. Taylor Fravel, "Dangerous Confidence? Chinese Views on Nuclear Escalation," *International Security* 44, no. 2 (October 1, 2019): 61–109, https://doi.org/10.1162/isec_a_00359.

Conclusion

This report outlined a high-level strategic approach to nuclear risk reduction. It is not, however, the final word on this issue and is intended mainly as a first attempt to consider the problem. We hope the report will help guide rapid action in light of recent philanthropic shortfalls. **We encourage readers to reach out about any potential mistakes in the analysis of this report.**

To summarize the key points of this report, funders face extreme uncertainty when thinking about nuclear risk and risk reduction measures. Nonetheless, several simple principles can help guide action-oriented philanthropists:

- Prioritize minimizing expected war damage, not possible intermediate goals, such as disarmament for disarmament's sake.
- Recognize that large all-out wars may be disproportionately worse than limited nuclear use.
- Recognize that there exists extreme uncertainty about different interventions, and that accidental nuclear use cannot be ruled out.
- Leverage the power of "impact multipliers," including focusing on neglected strategies, prioritizing the "great powers," and leveraging societal resources.
- Use these insights to identify and rank the most promising interventions and re-shape the field of nuclear security for the better in light of the recent funding shortfalls.

Addressing the challenges outlined in this report will require large-scale philanthropic investment and innovative thinking. The world is entering a new and potentially dangerous nuclear age, and society is vastly under-investing in mitigating this global catastrophic risk.

Appendix 1: Probabilities of Nuclear Catastrophe

The following table is a partial repository of estimates of the probability of nuclear war, including a calculation of the geometric mean of odds. We encourage readers to create copies and add to this table.

The **Aggregated Results** section presents headline figures: first, the arithmetic mean of probabilities (see Rodriguez, 2019) of a nuclear war occurring and the geometric mean of odds of a nuclear war occurring. Next, it presents crowd estimates derived from Metaculus probabilistic forecasting tournaments for (1) the probability that a nuclear war occurs and kills between 10% and 95% of the human population — an unprecedented global catastrophe — and (2) the probability that a nuclear war occurs and kills more than 95% of the population — an extinction or near-extinction scenario.

The section on **Probabilities of Nuclear War** collects several point estimates of the probabilities of nuclear wars, aggregated in part from Luisa Rodriguez's 2019 analysis.

Probabilities of Catastrophe and Near-Extinction uses estimates from a Metaculus forecasting tournament to derive conditional probabilities of global catastrophe (>10% of the human population dying) and near-extinction (>95% of the human population dying) due to nuclear use.

P(nuclear war), annualized annualizes the probabilities given in other sections and calculates the arithmetic mean of probabilities.

Odds of nuclear war converts probabilities to odds, calculates the geometric mean of odds, and converts the result back into a probability.

Variable	Value	Source	Notes
Aggregated Results			**This section provides headline figures for the calculations performed below.**
Arithmetic mean of probabilities (nuclear war occurs)	1.22%	Calculation	This cell estimates the probability of outbreak of nuclear war in a given year using the arithmetic mean of probabilities of

			various annualized point estimates (see Rodriguez, 2019 for the basis of these calculations)
Geometric mean of odds as probability (nuclear war occurs)	0.986%	Calculation	This cell estimates the probability of outbreak of nuclear war in a given year using the arithmetic mean of probabilities. This may be preferred: see "When Pooling Forecasts, Use Geometric Mean of Odds"
Crowd Estimate (10%<pop. decline<95%)	0.15%	Calculation based on crowdsourced forecast (Metaculus)	This cell uses a crowdsourced forecast (May 2023) to estimate the probability that a nuclear war occurs and kills between 10% and 95% of the human population.
Crowd Estimate (pop. decline>95%)	0.01%	Calculation based on crowdsourced forecast (Metaculus)	This cell uses a crowdsourced forecast (May 2023) to estimate the probability that a nuclear war occurs and kills over 95% of the human population.
Probabilities of Nuclear War			**This section collects various estimates of the probability of nuclear war**
Applied Physics Laboratory (2021): Annual Probability of Nuclear War (best guess)	0.010000	Applied Physics Laboratory, "On Assessing the Risk of Nuclear War"; See also article in *Bulletin of Atomic Scientists* for same discussion	This is an order-of-magnitude estimate of the annual probability of nuclear war based on reasoning about past crises.
Applied Physics Laboratory (2021) Annual Probability of Nuclear War (low)	0.003000	Applied Physics Laboratory, "On Assessing the Risk of Nuclear War"	S/A
Applied Physics	0.030000	Applied Physics	S/A

Laboratory (2021) Annual Probability of Nuclear War (high)		Laboratory, "On Assessing the Risk of Nuclear War"	
P(Nuclear War), historical frequency	0.012987	Based on one use since 1945	This estimate uses the fact that there has been one use of nuclear weapons since 1945 to create a naïve base rate. This estimate is slightly lower than Rodriguez, 2019, given the additional four years that have passed.
Lugar (2005): Annual P(Nuclear War)	0.022100	Expert Survey: The Lugar Survey on proliferation threats and responses.	From Rodriguez, 2019.
Sandberg&Bostrom (2008): P(Nuclear War kills at least 1 million)	0.003900	Expert Survey: Sandberg and Bostrom, 2008.	From Rodriguez, 2019.
P(nuclear detonation by state actor causes at least 1 fatality)	0.004000	Unpublished Good Judgment Data from Rodriguez, 2019.	From Rodriguez, 2019.
P(fatality due to deliberate nuclear detonation by 2024)	0.040000	Metaculus Nuclear Risk Tournament	This estimate is based on a crowdsourced forecast.
Probabilities of Catastrophe and Near-Extinction			
P(Global Pop. Decline >10%)	38%	By 2100, will the human population decrease by at least 10% during any period of 5 years?	This row estimates the probability that the human population will decrease by at least 10% during any period of 5 years between now and the end of this century. The estimate is based on a crowdsourced probabilistic forecast on the forecasting platform Metaculus as of May 2023. Like other estimates in this spreadsheet, it is

			highly subjective.
P(GCR caused by nuclear war | GCR occurrence)	31%	If a global catastrophe happens before 2100, will it be principally due to the deployment of nuclear weapons?	This row estimates the probability that any such catastrophe, if it occurs, will be "principally due to the deployment of nuclear weapons."
P(nuclear war causes pop. decline >10%)	12%	Calculation	This row multiplies the probability of a 10% population decline and the probability that such a decline will be due to nuclear weapons to calculate a conditional probability that nuclear war causes a population decline of at least 10% by 2100.
Years to 2100	77.60	Input	This row counts the years left until the end of 2100, the resolution date of the forecasts described above.
P(nuclear GCR), annualized	0.16%	Calculation	This row annualizes the probability that a nuclear war causes a population decline of at least 10%.
P(nuclear catastrophe causes near-extinction | nuclear catastrophe)	5.00%	If a nuclear catastrophe occurs that results in a reduction of the global population of at least 10% by 2100, will the global population decline by more than 95% relative to the pre-catastrophe population?	This row estimates the probability that, conditional on a 10% population decline, there will be a greater than 95% population decline by 2100. Like several of the estimates above, it is based on a crowdsourced probabilistic forecast as of May 2023.
P(10%<fatalities<95% | nuclear catastrophe)	95.00%	Calculation	This row is simply calculated from the preceding row; it is the probability that, conditional on a 10% population decline, this

			decline will be between 10% and 95%
P(nuclear GCR kills <95%)	0.15%	Calculation	This row multiplies the annual probability that nuclear catastrophe (>10% population decline) occurs and the probability that such a catastrophe kills no more than 95% of the population.
P(nuclear GCR kills >95%)	0.01%	Calculation	This is the annual probability that a nuclear catastrophe occurs and kills more than 95% of the human population, calculated from the estimate above.
P(nuclear war), annualized			
Applied Physics Laboratory best guess	0.010000	See row 10	The estimates in "On Assessing the Risk of Nuclear War" are already annualized.
Naive historical base rate	0.012987	See row 13	S/A
Lugar expert survey	0.022100	See row 13	S/A
GCR Survey	0.003900	See row 15	I think this is likely a slight underestimate (because there are nuclear wars with <1M fatalities), but the Metaculus question on nuclear detonations is likely a slight overestimate (because it likely includes acts of terrorism, etc.)
Good Judgment	0.004000	See row 16	
Metaculus (nuclear detonation)	0.02020410289	See row 17	Annualizing the estimate from row 17.
Metaculus (nuclear GCR)	0.00161	See row 24	
Arithmetic Mean of Probabilities	0.012199	Calculation	This row purports to calculate the arithmetic mean of probabilities of a

			nuclear war occurring. In fact, it aggregates several incommensurate estimates of different kinds of events (nuclear use of any kind, nuclear use that kills more than 1 million people, etc.). It does not include the estimate of a nuclear war occurring and killing more than 10% of the population, as this is a far less likely event.
Odds of nuclear war = P(nuclear war):P(not nuclear war)			This section uses the probabilities to create odds ratios, which are then used to calculate the geometric mean of odds.
Applied Physics Laboratory best guess	0.010101	Calculation	
Naive historical base rate	0.013158	Calculation	
Lugar expert survey	0.022599	Calculation	
GCR Survey	0.003915	Calculation	
Good Judgment	0.004016	Calculation	
Metaculus	0.020621	Calculation	
Geometric Mean of Odds	0.009956	Calculation	This row uses the odds in previous rows to calculate the geometric mean of odds, a measure used to aggregate probabilistic forecasts.
Geometric Mean of Odds (Converted back to probabilities)	0.009858	Calculation	This row converts the geometric mean of odds back to a probability.

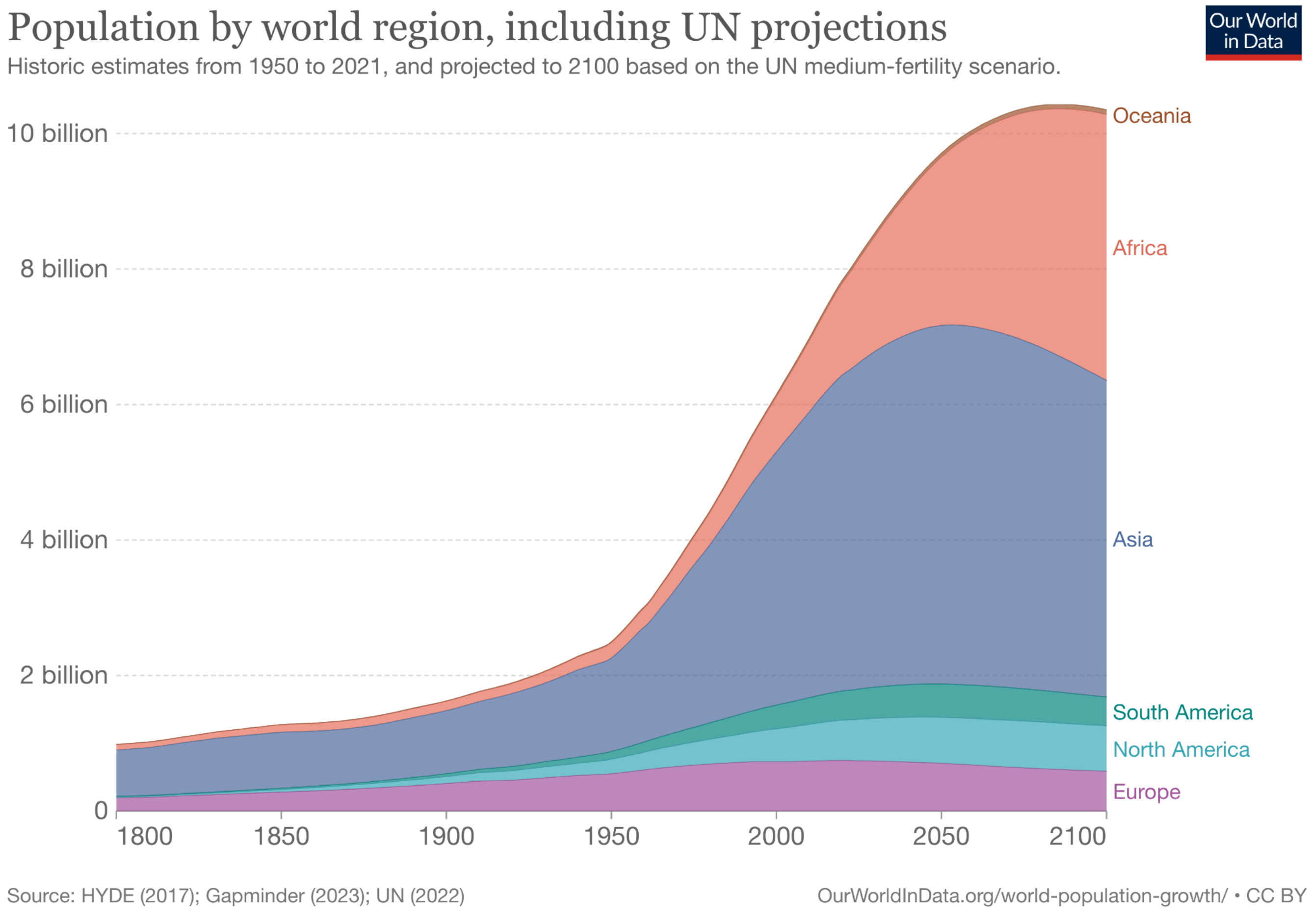
Appendix 2: Cost-Effectiveness and Break-Even Analyses

This section provides a rough attempt at modeling the cost-effectiveness of nuclear risk spending with a cost-effectiveness analysis (CEA) with point estimates. **This table serves as a first-draft proof-of-concept, and should not be taken literally, but as a starting point for more sophisticated models.** There are many serious shortcomings to this simplistic approach, but this kind of back-of-the-envelope calculation (BOTEC) helps to illustrate that nuclear risk reduction is plausibly highly cost-effective *even without accounting for long-term and second-order effects*. Carl Shulman and Elliott Thornley (building on earlier work) have recently made the case for simple cost-benefit analyses of global catastrophic risks in public policy, writing "Governments should be spending much more on averting threats from nuclear war, engineered pandemics, and AI. This conclusion follows from standard cost benefit analysis. We need not assume longtermism, or even that future generations matter."[249] The table below attempts to extend a similar analysis to philanthropic interventions. It is not possible to represent the full horrors of nuclear war with such models, but they may nonetheless help us to prioritize the distribution of limited societal resources.

For simplicity, we use probability estimates from Metaculus for "Global Catastrophe" and "Near-Extinction," as these are some of the only available estimates that allow us to derive escalation probabilities and distinguish between the largest nuclear wars and *any* nuclear use. Similar calculations could be performed using the geometric mean of odds obtained above.

Note that world population is expected to level off around 10 billion this century; the cost-effectiveness analysis simplistically assumes a population of 9 billion. If population were to continue growing, the costs would be far higher:

[249] Carl Shulman and Elliott Thornley, "How Much Should Governments Pay to Prevent Catastrophes? Longtermism's Limited Role," https://philpapers.org/archive/SHUHMS.pdf



Source: Our World in Data.

Important Omitted Effects of Nuclear War

This BOTEC is designed for simplicity, and therefore leaves out many important considerations, including:

- **Non-fatal casualties** — this spreadsheet only counts fatalities, when the effects of nuclear war include flash burns, injuries from the blast and falling debris, and acute radiation sickness, among others.[250] These injuries are not included in the estimate, and their effect on health and wellbeing may be large;
- **Long-term radiation damage** — Moreover, the effects of radiation are not limited to acute radiation sickness, but include increased risk of cancer, as demonstrated by

[250] For a detailed overview of these effects, see OTA, *The Effects of Nuclear War*.

the cohort study of the survivors of the atomic bombings of Japan.[251] (As far as we know, there were no hereditary effects of these bombings, so the long-term radiation effects would be limited to one generation in an expanded model.[252]) These effects are not modeled.

- **Excess deaths and increased pandemic risks** — Widespread damage to critical infrastructure, including damage to roads from the blast and to communications via electromagnetic pulse, and the potential that hospital intensive care units would be overwhelmed, may increase excess deaths among people who would otherwise seek medical care.[253] Moreover, some public health experts believe that weakened immune systems and the post-war environment could increase the risk of a catastrophic pandemic. This is not included in this model.
- **Economic effects** — This model (unlike e.g. some of Founders Pledge's climate modeling), does not include the economic effects of a nuclear war, even though those effects would likely be severe (including halting air travel, destroying critical infrastructure, damaging information and communications technology, and more). Even small disruptions can have large economic effects; the temporary restrictions to European airspace following the magnitude 4 eruption of Eyjafjallajökull, for instance, are estimated to have caused $5 billion in losses to the global economy, and the September 11 attacks on the United States cost around $60 billion, by some estimates.[254]
- **Animal suffering** — This model focuses only on human deaths, and excludes the suffering of animals that would be caused by nuclear war and nuclear climate effects;
- **Grief, PTSD, and mental health** — Moreover, the model does not include the subjective wellbeing effects of survivors' grief, post-traumatic stress disorder, and overall mental health damage of a nuclear war. These may include social stigma from

---

[251] Kenji Kamiya et al., "Long-Term Effects of Radiation Exposure on Health," *The Lancet* 386, no. 9992 (August 1, 2015): 469–78, https://doi.org/10.1016/S0140-6736(15)61167-9.
[252] Ibid.
[253] Notably, these kinds of risks have received little more than speculative attention from nuclear scholars. As James Scouras writes, "the physical consequences to the infrastructures that sustain societies – power, water, finance, transportation, etc. – has never been a focus of nuclear weapons effects research," and the effect of their loss on overall mortality is highly uncertain. James Scouras, "Nuclear War as a Global Catastrophic Risk," *Journal of Benefit-Cost Analysis* 10, no. 2 (2019): 274–95, https://doi.org/10.1017/bca.2019.16.
[254] Lara Mani, Asaf Tzachor, and Paul Cole, "Global Catastrophic Risk from Lower Magnitude Volcanic Eruptions," *Nature Communications* 12, no. 1 (December 2021): 4756, https://doi.org/10.1038/s41467-021-25021-8; S. Brock Blomberg and Gregory D. Hess, "Estimating the Macroeconomic Consequence of 9/11," *Peace Economics, Peace Science and Public Policy* 15, no. 2 (January 6, 2009), https://doi.org/10.2202/1554-8597.1167.

having been exposed to radiation, a phenomenon that has been widely observed in many nuclear incidents, where survivors become social outcasts.[255]

- **Trajectory changes** – Like great power war more broadly, nuclear war could change the trajectory of human civilization (for example, by speeding up the unsafe development of dangerous technology, or by locking in authoritarian values in a post-war world). These effects could be immense, but they are not modeled here.
- **Increased likelihood of other WMD use** – Even a small nuclear war could precipitate the use or development of other weapons of mass destruction, like strategic biological weapons;
- **Breaking "nuclear taboo"** – More broadly, as this estimate is focused on a single instance of nuclear use, it does not engage with the idea that any nuclear weapons use might break the "nuclear taboo," potentially making future nuclear weapons use more likely (and more broadly, undermining the international order). As discussed above, it is not obvious, however, that a breach of the taboo would necessarily weaken it; indeed, some scholars suggest that occasional norm violations can help stimulate opprobrium and even strengthen a norm against nuclear use.[256]

Annotated Calculations

***NB:** The linked spreadsheet contains hidden rows for "optional subjective input," allowing the reader to experiment and substitute different values for key inputs to the calculations.*

Variable	Value	Source	Notes
Annualized Probabilities of Nuclear Catastrophe			
P(nuclear GCR, 10%<fatalities<95%)	0.153%	P(nuclear catastrophe)	This row draws from the P(nuclear catastrophe) spreadsheet an annualized probability that a nuclear catastrophe occurs and kills between 10% and 95% of the human population

[255] "If history is any guide, some degree of social stigma would also attach to individuals from areas associated with radiation, as it did in the wake of Hiroshima and Nagasaki, Chernobyl, Goiânia, and the 1999 nuclear criticality accident in Tokaimura, Japan." James Scouras, "On Assessing the Risk of Nuclear War."
[256] Scouras, "Nuclear War as a Global Catastrophic Risk," 283.

P(nuclear near-x)	0.008%	P(nuclear catastrophe)	This row draws from the P(nuclear catastrophe) spreadsheet an annualized probability that a nuclear catastrophe occurs and kills 95% of the human population
P(nuclear GCR, total)	0.161%	Calculation	This row adds the two probabilities for the total annual probability of a global catastrophic nuclear event that kills more than 10% of the population.
Effects of Catastrophe			
Nuclear GCR expected fatalities	900,000,000	Nuclear Fatalities	This row estimates the mean deaths from a nuclear catastrophe that kills 10% of the human population, simplistically assuming a uniform distribution of fatalities and a world population of 9 billion humans. NB that 10% of the population is on the low end of what the crowdsourced forecasting platform asked about; this assumes that the deaths cluster around the low end. A more sophisticated analysis would use probability distributions.
Nuclear near-extinction expected fatalities	8,550,000,000	Nuclear Fatalities	This row estimates the mean deaths from a nuclear catastrophe that kills more than 95% of the human population
Annualized Expected Cost(nuclear catastrophe)			

EC(nuclear GCR), conservative estimate	1,379,847.99	Calculation	This row calculates the expected cost (EC = probability*consequence) of a nuclear catastrophe that kills 10% of the human population. NB. this is a conservative estimate, because it rounds the population loss between 10% and 95% to a population loss of 10%
EC(nuclear near-x)	689,924.00	Calculation	This row calculates the expected cost of a nuclear catastrophe that kills over 95% of the human population. Like the previous row, it is a conservative estimate, because it assumes 95% exactly, not >95%.
EC(total nuclear catastrophe)	2,069,771.99	Calculation	This row calculates the total expected cost of a nuclear catastrophe by adding the previous two rows together.
Doubling risk reduction spending			
Cost to double private philanthropic effort	$32,100,000	Candid Foundation Maps (2021)	The Peace and Security Funding Index assumes that $32.1 Million (total grants) were spent on nuclear security in 2021. Note that this may be significantly lower in the near future. We have chosen 2021 because it is the most recent year for which the data are roughly accurate.
Relative risk reduction from doubling total spending	1%	Subjective Input	This is a subjective input for the relative risk reduction achieved from a doubling of total

			philanthropic spending. The entire calculation hinges on this highly subjective input. As a point estimate, it obscures much complexity, including questions of marginal cost-effectiveness declines.
EC(nuclear catastrophe | doubling intervention)			
P(nuclear GCR | no intervention)	0.15332%	See above	This row represents the probability of a 10% population decline due to nuclear war with business-as-usual.
P(nuclear-near X | no intervention)	0.00807%	See above	This row represents the probability of a 95% population decline due to nuclear war with business-as-usual.
P(nuclear GCR | intervention)	0.15178%	See above	This row calculates the effect of a 1% relative risk reduction in the probability of 10% population decline due to nuclear war, after a doubling of philanthropic investment. It assumes that the risk reduction is equal on GCR and X-risks
P(nuclear-near X | intervention)	0.00799%	Calculation	This row calculates the effect of a 1% relative risk reduction in the probability of 95% population decline due to nuclear war, after a doubling of philanthropic investment. It assumes that the risk reduction is equal on GCR and X-risks

EC(nuclear GCR | doubling intervention)	1,366,050	Calculation	This row uses the reduced probability of nuclear war to calculate a new expected cost of nuclear catastrophe that kills 10% of the population, conditional on doubled philanthropic investment.
EC(nuclear near-X | doubling intervention)	683,025	Calculation	This row uses the reduced probability of nuclear war to calculate a new annual expected cost of nuclear near-extinction that kills 95% of the population, conditional on doubled philanthropic investment.
EC(nuclear catastrophe, total | doubling intervention)	2,049,074	Calculation	This row uses the reduced probability of nuclear war to calculate a new expected total cost of nuclear catastrophe in lives lost.
CEA, Cost per life saved			
Expected lives saved (nuclear GCR)	13,798.48	Calculation	This row calculates the difference in expected cost before and after the intervention on a 10% population decline event.
Expected lives saved (nuclear near-X)	6,899.24	Calculation	This row calculates the difference in expected cost before and after the intervention on a 95% population decline event.
Expected lives saved (total nuclear catastrophe)	20,697.72	Calculation	This row calculates the total difference in expected cost before and after the intervention on global nuclear catastrophe.
Cost per life saved (total nuclear catastrophe)	$1,551	Calculation	This row divides the cost of the intervention by the total number of lives

			saved to estimate cost-per-life-saved.
CEA (WELLBY)			
Probability that death is under 5	6.82%	Calculation	This row uses the UN's tables and projections for population by age to estimate the probability that a catastrophe (randomly distributed over the next 100 years) is likely to affect a person under 5 years old. This matters if we place different moral weights on infant deaths.
Probability that death is over 5	93.18%	Calculation	This row in turn estimates the probability that a catastrophe is likely to affect a person older than 5 years.
WELLBY-equivalent of 1 death under 5	244.8	Founders Pledge Moral Weights	This row provides the WELLBY-equivalent value of averting the death of a human under 5 years old.
WELLBY-equivalent of 1 death over 5	166.9	Founders Pledge Moral Weights	This row provides the WELLBY-equivalent value of averting the death of a human over 5 years old.
WELLBY loss averted through intervention	3,564,349.55	Calculation	This row calculates the annual expected WELLBY loss averted through this intervention.
Cost per WELLBY	$9	Calculation	This row estimates the cost-per-WELLBY of the intervention by dividing the cost of doubling philanthropic investment by the benefit of averting WELLBY loss.
Multiples of AMF	**2.5**	Calculation	Assuming $3,845

			cost-per-life-saved for AMF
Multiples of GiveDirectly	**18**	Calculation	Assuming $166 cost to double income
Breakeven Analysis (GiveDirectly bar)			
Breakeven cost-per-WELLBY	$166.00	Moral Weights and Comparisons	This row gives the "breakeven cost" per WELLBY — the cost-effectiveness bar that a hypothetical intervention must meet in order to be more cost-effective than direct cash transfers to the world's poorest people, a common benchmark in charity evaluation.
WELLBY loss averted to break even	193,373.49	Calculation	WELLBYs saved = (cost to double effort)/(cost-per-WELLBY)
Lives saved to break even	1,122.90	Calculation	Lives saved to break even = (WELLBY Loss averted to break even)/((P(death is under 5)*(WELLBY-equivalent of death under 5)+(P(death is over 5)*(WELLBY-equivalent of death over 5))
EC(nuclear catastrophe, total | doubling intervention), to break even	2,068,649.09	Calculation	EC(nuclear catastrophe, total | intervention) = EC(nuclear catastrophe | no intervention) - (lives saved to break even)
Implied relative risk reduction to break even	**0.05%**	Calculation	This row calculates the relative risk reduction of a hypothetical intervention of the magnitude described above required to break even with direct cash transfers.

Appendix 3: Nuclear Winter Studies

[This table was generated by a human-machine team (Christian Ruhl and GPT4).]

Study	Publication Year	Warhead Number, Yield (where known)	Estimated Cooling (temperature)	Cooling Length (Years)	Notes and Study Quality
Crutzen and Birks, "The Atmosphere after a Nuclear War: Twilight at Noon," *Ambio*.	1982	Not specified	Severe cooling	Not specified	Early study, no specific war scenario.
Turco et al., "Nuclear Winter: Global Consequences of Multiple Nuclear Explosions," *Science*.	1983	5000MT delivered by **10,400** warheads	**-15º to -25º**	**1+** years	Early study introducing the idea of nuclear winter and global climate effects.
Turco et. al., "Climate and Smoke: An Appraisal of Nuclear Winter," *Science.*	1990	Not specified. Assume arsenal at 25,000 warheads carrying 10,000 megatons	**-10º to -20º** midsummer cooling in northern mid-latitudes. As high as **-35º** cooling locally	**1-3** years	Summary of research since 1983. Low transparency on modeling, inputs, etc.
[no publications]	1990-2003	-	-	-	Apparent "research winter"
Robock et al., "Climatic consequences of regional nuclear conflicts," *Atmospheric Chemistry and Physics.*	2007	**100** warheads (15 kt each)	**-1.25º** global average cooling	**3-10+** years (see fig. 3 on p. 2005)	Updated climate model on smaller conflict. See criticisms of study methodology in Reisner et al. (2018).

Robock et al., "Nuclear winter revisited with a modern climate model and current nuclear arsenals: Still catastrophic consequences," *Journal of Geophysical Research*	2007	Two scenarios: 1. 150Tg, entire arsenal (authors estimate **~21,000** warheads) 2. 50Tg, 1/3 of arsenal (i.e. **~7,000** warheads)	150Tg case: **-7º to -8º** (i.e., "Ice Age") 50Tg case: **-3º to -4º**	**10+ years** (see fig. 2 on p. 5)	Low transparency on results (e.g. do not explicitly state 50Tg global mean temperature decrease). Authors assume that disarmament is the only solution: " only nuclear disarmament will completely remove the possibility of a nuclear environmental catastrophe."
Mills et al., "Multidecadal global cooling and unprecedented ozone loss following a regional nuclear conflict," *Earth's Future*	2014	**100** warheads (15 kt each)	**-1.1º to -1.6º**	**25+ years**	Very similar to Robock et al. (2007a). Political call to action: "Knowledge of the impacts of 100 small nuclear weapons should motivate the elimination of more than 17,000 nuclear weapons that exist today"
Pausata et al., "Climate effects of a hypothetical regional nuclear war: Sensitivity to emission duration and particle composition," *Earth's Future*	2016	**100** warheads (15kt each)	**-0.12ºC to -1.3ºC in the first year, depending on emission time and scenario**	**10+ years. Some cooling present after 20 years**	The main difference between this and the other estimates of India-Pakistan war with the 5Tg assumption is the injection time period. Research supported by Swedish Physicians against Nuclear Weapons.
Reisner et al., "Climate Impact of a Regional Nuclear Weapons Exchange: An Improved Assessment	2018	**100** warheads (15 kt each)	**-0º to 0.5º**, effects mostly limited to polar regions (i.e. not affecting most human activity)	**<5 years**	Model of a "regional" India-Pakistan exchange, testing the Robock et al. (2007) claim that a regional nuclear war would lead to significant global cooling. Study authors affiliated with U.S.

Based On Detailed Source Calculations" *JGR Atmospheres*					National Laboratories.
Xia et al. "Global food insecurity and famine from reduced crop, marine fishery and livestock production due to climate disruption from nuclear war soot injection," *Nature Food*	2022	**100-4,000 warheads**: • 100 15kt warheads • 250 15kt warheads • 250 50kt warheads • 250 100kt warheads • 500 100kt warheads • 4,400 100kt warheads	**-2º to -16º**, depending on scenario (see figure 1, panel a)	**16+ years**, depending on scenario	NB apparent fatality estimates mean "at risk" of food insecurity: i.e. "could" in the following sentence is an expression of possibility, not probability: "We estimate more than 2 billion people could die from nuclear war between India and Pakistan, and more than 5 billion could die from a war between the United States and Russia" No adaptation: "we do not consider farm-management adaptations such as changes in cultivar selection, switching to more cold-tolerating crops or greenhouses and alternative food sources[...]."

About Founders Pledge

Founders Pledge is a global nonprofit empowering entrepreneurs to do the most good possible with their charitable giving. We equip members with everything needed to maximize their impact, from evidence-led research and advice on the world's most pressing problems, to comprehensive infrastructure for global grant-making, alongside opportunities to learn and connect. To date, they have pledged over $10 billion and donated more than $850 million globally. We're grateful to be funded by our members and other generous donors. founderspledge.com

About the Author

Christian Ruhl is a senior researcher at Founders Pledge and the fund manager for the Global Catastrophic Risks Fund. Christian's work focuses on understanding, forecasting, and mitigating global catastrophic risks, including risks from great power competition and weapons of mass destruction. Previously, Christian was the program manager for *The Future of the Global Order: Power, Technology, and Governance* at Perry World House, the University of Pennsylvania's global affairs think tank. After receiving his BA from Williams College, he studied on a Dr. Herchel Smith Fellowship at the University of Cambridge for two MPhil programs, one in History and Philosophy of Science and Medicine and one in International Relations and Politics, with dissertations on early modern state-sponsored science and Cold War strategy on tactical nuclear weapons and limited nuclear war. Christian was a member of the 2021 Project on Nuclear Issues Nuclear Scholars Initiative, serves on the External Advisory Board of the Berkeley Risk and Security Lab at the University of California, and supervises summer research projects for the Existential Risk Alliance. His writing has appeared in *The Atlantic*, *The Bulletin of the Atomic Scientists*, *Foreign Policy*, *Lawfare, Texas National Security Review,* and more.