

SPECIAL ISSUE ON LONG-TERM RISKS AND FUTURE GENERATIONS

Existential Terrorism: Can Terrorists Destroy Humanity?

Zachary Kallenborn^{1,2,3}  and Gary Ackerman⁴

¹Strategic Technologies Program, Center for Strategic and International Studies (CSIS), Washington, DC, USA, ²Schar School of Policy and Government, George Mason University, Fairfax, VA, USA, ³Unconventional Weapons and Technology Division, START, University of Maryland, College Park, MD, USA and ⁴College of Emergency Preparedness, Homeland Security and Cybersecurity, University at Albany, Albany, NY, USA

Corresponding author: Zachary Kallenborn; Email: zkallenborn@gmail.com

Abstract

Mass-casualty terrorism and terrorism involving unconventional weapons have received extensive academic and policy attention, yet few academics have considered the broader question of whether such behaviours could pose a plausible risk to humanity's survival or continued flourishing. Despite several terrorist and other violent non-state actors having evinced an interest in causing existential harm to humanity, their ambition has historically vastly outweighed their capability. Nonetheless, three pathways to existential harm exist: existential attack, existential spoilers and systemic harm. Each pathway varies in its risk dynamics considerably. Although an existential attack is plausible, it would require extraordinary levels of terrorist capability. Conversely, modest terrorist capabilities might be sufficient to spoil risk mitigation measures or cause systemic harm, but such actions would only result in existential harm under highly contingent circumstances. Overall, we conclude that the likelihood of terrorism causing existential harm is extremely low, at least in the near to medium term, but it is theoretically possible for terrorists to intentionally destroy humanity.

Keywords: Existential risk; risk analysis; terrorism

I. Introduction

On occasion, terrorist groups have sought to destroy human civilisation. The apocalyptic cult Aum Shinrikyo carried out terror attacks with the goal of inciting an apocalyptic war. Originally a religious organisation, Aum Shinrikyo's leader, Asahara Shoko, became deeply enmeshed in apocalyptic literature.¹ He came to believe that Aum must incite the coming apocalypse, with himself emerging as a New Christ. After failing to win election to the Japanese Diet, Aum Shinrikyo turned to violence, seeking chemical, biological and nuclear weapons and even an earthquake-making device.² Then, on 20 March 1995, Aum Shinrikyo

¹ R Danzig et al, "Aum Shinrikyo: Insights Into How Terrorists Develop Biological and Chemical Weapons, 2nd ed." (*Center for a New American Security*, December 2012) <https://s3.us-east-1.amazonaws.com/files.cnas.org/hero/documents/CNAS_AumShinrikyo_SecondEdition_English.pdf?mtime=20160906080510&focal=none> (last accessed 3 April 2023).

² RJ Lifton, *Destroying the World to Save It: Aum Shinrikyo, Apocalyptic Violence, and the New Global Terrorism* (New York, Metropolitan Books 1999) pp 119–20.

conducted a sarin attack on the Tokyo subway system, killing thirteen and injuring thousands more.³

Aum Shrinkyo and similar maximalist terrorist groups constitute a small and distinct subset of terrorists, those who seek to perpetrate what we label “existential terrorism”. By this, we mean that the terrorist actor aims to cause the extinction of humanity or at least cause a global collapse of civilisation (or come so close to either that it does not matter).⁴ Existential terrorism represents an extreme form of terrorism because the consequences are many orders of magnitude higher than even the largest mass-casualty terrorist attacks perpetrated thus far. Consequently, the capabilities and motivations involved in conducting such an attack are quite different from most other terrorist actions.

Put simply, existential terrorism is action by a terrorist actor (non-state group or even individual) that intentionally or unintentionally might result in existential harm to humanity. Yet even this simple formulation poses theoretical complications. First, existential terrorism as defined may not meet the strict criteria for terrorism put forward by many scholars because it need not entail coercion, intimidation or sending a message to a broader audience.⁵ After all, if every human is dead, who would hear the message? It is important to note that it is not whether a message is received that is important, but rather that the perpetrator intends to send one. Perhaps the terrorist is sending a message to God, nature, a future Earth species or an alien race. Similarly, an attack to galvanise others to act, such as Aum’s attempt to incite a cataclysmic global war, involves sending messages, albeit tacit ones. In any event, it is intuitively problematic to ignore the most extreme possibilities of non-state violence merely because they do not strictly satisfy the messaging criterion of some definitions of terrorism. Therefore, in the current paper we include any extremist violence by any non-state actor, even if it would technically fall into the category of mass murder as opposed to terrorism narrowly defined.

Existential terrorism is related to but distinct from the more commonly studied topics of mass-casualty and chemical, biological, radiological and nuclear (CBRN) terrorism. Most forms of mass-casualty and CBRN terrorism will not cause existential harm, even in extraordinary cases. If a terrorist detonated a nuclear weapon in New York City, the attack certainly would be catastrophic (especially for New Yorkers), but it would not be globally existential, except perhaps as a third- or fourth-order consequence. Nonetheless, mass-casualty and CBRN terrorism are still relevant to existential risk because at least biological and nuclear weapons have the potential to bring about existential harm. Moreover, any form of extreme terrorism attack could distract states from mitigating existential harms, a topic we will return to later.

Like many, we approach the notion that terrorists might succeed in destroying humanity in the foreseeable future with a healthy dose of scepticism. Indeed, most other discussions of existential risk involve either large-scale natural phenomena or the actions of substantial polities and their attendant infrastructures. Yet, thinking through the risk possibilities of existential terrorism is worthwhile for at least two reasons. First, it allows

³ “Japan Marks 25th Anniversary of Deadly Tokyo Subway Attack by Cult” (*Kyodo News*, 20 March 2020) <<https://english.kyodonews.net/news/2020/03/c97110923597-update1-japan-marks-25th-anniversary-of-deadly-tokyo-subway-attack-by-cult.html>> (last accessed 14 June 2023).

⁴ This breakdown for existential risk is found in N Bostrom, “Existential risk prevention as global priority” (2013) 4(1) *Global Policy* 15, and echoed in T Ord, *The Precipice: Existential Risk and the Future of Humanity* (New York, Hachette Books 2020). There is a third type of existential risk adopted by these authors, namely locking the world into a totalitarian future, but this is not accepted by all scholars and so is dealt with separately in this paper (Box 1).

⁵ START’s Global Terrorism Database does not strictly require evidence of coercion or attempts to send a message. However, the FBI’s and many other definitions include an attempt to intimidate or coerce. For a list of 250-plus academic, governmental and intergovernmental definitions of terrorism, see JJ Easson and AP Schmid, *Routledge Handbook of Terrorism Research* (Abingdon, Routledge 2011) pp 99–157.

us to examine the lower bound for the size of a group of malefactors who could bring about the end of humankind or, conversely, the upper bound of the scale of harm that an individual or small group could conceivably inflict. Second, the admittedly low probability of terrorists acquiring the capability to inflict existential harm forces us to think beyond the obvious direct “attacking” modalities usually associated with terrorist actions to more nuanced considerations of their broader potential for harm. Therefore, even if we were to conclude that existential terrorism is infeasible for now, the very act of considering how such terrorism could manifest can be instructive to the larger topics of terrorism and existential risk assessment.

II. Existential terrorist motivations

A central question is why terrorists or other non-state actors might seek to destroy humanity. If there are indeed few or no feasible motivations for doing so, then the threat is diminished irrespective of any capability to act upon it. After all, if, as many claim, terrorists are political actors who are rational within their frame of reference,⁶ what conceivable reason could they have for destroying themselves, their putative constituents and, indeed, the entire society of human beings whom they seek to influence? What cause could destroying the world possibly serve? Unfortunately, there are plausible conditions under which individuals and groups – both traditional terrorists and otherwise – might conclude that bringing about existential levels of harm is a necessary or preferable path to pursue.

Beginning with individuals, the sheer diversity of human thought and beliefs, together with the law of large numbers, is enough to create a *prima facie* argument that at least some individuals will possess world-ending or civilisation-ending motivations. Out of roughly 8 billion human beings on the planet, of whom a not-insignificant number suffer from some form of mental illness, emotional trauma or solipsism, it is almost certain that at least a handful will decide that exterminating humanity is the answer to their perceived problems or ambitions. Some of the many ways that this might occur could include: extreme depression or personal trauma leading to acute misanthropy in a macroscopic version of murder-suicide; individuals suffering from severe mental illness that receive delusional instructions to end the world; and a host of idiosyncratic belief systems that might appear irrational to outsiders but are internally logically consistent in leading to a desire for Armageddon. As Stanford University epidemiologist Stephen Luby quips, “There’s no shortage of sociopaths.”⁷

This does not paint a very promising initial picture with respect to motivations, for as Sir Martin Rees puts it, “If there were millions of independent fingers on the button of a Doomsday machine, then one person’s act of irrationality, or even one person’s error, could do us all in.”⁸ One might then jump to the conclusion that exploring the motivational side of the threat is a waste of time, since we should just regard the presence of world-destroying motivation as axiomatic. Fortunately, there is a silver lining, and a very important one at that. This arises from the observation that, although it is conceivable that

⁶ M Crenshaw, “The logic of terrorism: terrorist behavior as a product of strategic choice” in W Reich (ed.), *Origins of Terrorism: Psychologies, Ideologies, Theologies, States of Mind* (Cambridge, Cambridge University Press 1990) pp 247–60.

⁷ Quoted in J Berger, “What’s Likely to Cause Human Extinction – and How Can We Avoid It?” (*Stanford Earth Matters*, 19 February 2019) <<https://earth.stanford.edu/news/whats-likely-cause-human-extinction-and-how-can-we-avoid-it>> (last accessed 15 February 2023).

⁸ M Rees, *Our Final Hour: A Scientist’s Warning. How Terror, Error, and Environmental Disaster Threaten Humankind’s Future in This Century – on Earth and Beyond* (New York, Basic Books 2003).

a so-called “superempowered individual”⁹ could gain the capability to inflict existential-level harm, it is much more difficult for an individual to currently do so along any of the pathways described below than for a group of at least several people acting in concert. And if a group needs to act together, in most cases its members need to all be similarly motivated. This brings to the forefront the matter of ideology, since it is assumed that few if any of the other pathways to radicalisation like familial obligation, romantic entanglement or thrill-seeking¹⁰ would result in the desire for or acquiescence to worldwide destruction. And as systems of prescriptive beliefs designed to be shared,¹¹ there is likely to be a far smaller set of ideologies that might motivate groups to bring about a global extinction. Three of the most plausible candidate ideological referents are as follows, with most of them echoing long-standing tropes of apocalyptic millenarianism¹²:

- *Cathartic utopianism*: Groups who fall into this category generally believe that the current corrupt or corrupted world must be destroyed in order to usher in a new, better one. Such eschatological beliefs can arise in a variety of ideologies, including the more extreme interpretations of most major religious traditions (such as apocalyptic strains within far-right Christianity¹³ and both Sunni and Shi’i jihadism¹⁴), so-called new religious movements or “cults”, which are often syncretic and practice social coercion, and certain UFO-inspired groups who seek salvation from extra-terrestrial beings. Aum Shinrikyo; the Covenant, the Sword and the Arm of the Lord; and Heaven’s Gate are all notable prior examples of this type of group. Many adherents of these groups believe that a select few will survive the cataclysm to rebuild the new Eden, while others acknowledge that all physical life will be sacrificed but trust that immortal remnants will remain or transcend to new levels of existence. It is worth noting that it is only the more activist strain of cathartic utopianism, whose believers argue that they themselves must act to bring about the catharsis, which poses a danger, since the more passive forms are generally content to merely prepare themselves for the time when external forces will do the necessary destroying.
- *Extreme environmentalism*: This motivational impulse is grounded in the belief that it is humankind (or at least human civilisation) that must be extinguished, in order that non-human animals and the biosphere can survive and flourish. Such beliefs are nested in ideas of Deep Ecology,¹⁵ which rejects anthropocentrism; its ideologues often view non-human living entities and even non-living natural objects as having intrinsic value, sometimes equalling or exceeding that of human beings. Thus, one finds the Gaia Liberation Front, which argues in its manifesto that humankind is an “alien species”, “virus” or “cancer” that must

⁹ G Ackerman and L Pinson, “An Army of One: Assessing CBRN Pursuit and Use by Lone Wolves and Autonomous Cells” (2014) 26(1) *Terrorism and Political Violence* 226–45; T Friedman, *Longitudes and Attitudes* (New York, Farrar, Straus and Giroux 2002).

¹⁰ C McCauley and S Moskaleiko, “Mechanisms of Political Radicalization: Pathways Toward Terrorism” (2008) 20(3) *Terrorism and Political Violence* 415–33.

¹¹ G Ackerman and M Burnham, “Towards a Definition of Terrorist Ideology” (2019) 33(6) *Terrorism and Political Violence* 1160–90.

¹² K Umbrasas, “The Life Course of Apocalyptic Groups” (2018) 11(2) *Journal of Strategic Security* 32–53; N Cohn, *The Pursuit of the Millennium: Revolutionary Millenarians and Mystical Anarchists of the Middle Ages* (Oxford, Oxford University Press 1990).

¹³ See, eg, K Noble, *Tabernacle of Hate: Why They Bombed Oklahoma City* (Bristol, Voyageur 1998).

¹⁴ D Cook, *Contemporary Muslim Apocalyptic Literature* (Syracuse, Syracuse University Press 2005); T Furnish, *Holiest Wars: Islamic Mahdis, Their Jihads, and Osama bin Laden* (Westport, CT, Praeger 2005).

¹⁵ A Naess, “Spinoza and Ecology” (1977) 7 *Philosophia* 45–54.

be expunged from the planet.¹⁶ Indeed, in a 1989 article in *Earth First! Journal*, there was an “urgent” solicitation for “scientific research on a species specific virus that will eliminate *Homo shiticus* from the planet. Only an absolutely species specific virus should be set loose. Otherwise it will be just another technological fix. Remember, Equal Rights for All Other Species.”¹⁷ Or the influential Finnish activist Pentti Linkola declaring that, “if there were a button I could press, I would sacrifice myself without hesitating, if it meant millions of people would die”.¹⁸ Not all of this ilk want to necessarily wipe out all human life; some who fall under the rubric of anarcho-primitivists or neo-Luddites like Ted Kaczynski (the Unabomber) would be happy to merely destroy civilisation and send humanity back to the Stone Age.¹⁹

- *Strong negative utilitarianism*²⁰: Within moral philosophy, there is a variant of traditional utilitarianism that focuses on minimising the pain and suffering (as opposed to maximising the happiness) of as many people as possible. The “strong” version of this negative utilitarian worldview could logically lead to what R.N. Smart described in 1958 as “the benevolent world-exploder”,²¹ where, to reduce the greatest amount of suffering for the greatest number of people, it becomes an act of altruism to end all human life. While this view has elicited many detractors among philosophers and we have not yet seen any non-state actors who have explicitly adopted this as their platform, strong negative utilitarianism does at least provide a philosophical basis for an apparently “noble” motivation for inflicting existential harm. Moreover, the doctrines of many apocalyptic groups are redolent of the idea of “destroying the world to save it”,²² although there is generally a greater emphasis placed in these doctrines on the post-cataclysmic continuation or rebirth of humanity in either a material or immaterial form.

The advantage to understanding the possible motivations for terrorist and other violent non-state actors to cause existential harm is firstly a practical one: law enforcement and intelligence agencies can more easily identify those groups and individuals most likely to seek such outcomes and prioritise assessment of their capabilities and behaviour for indications that they are moving in this direction. Beyond merely facilitating the identification of possible threats, the type of motivation held by an actor can help us to anticipate the mechanisms by which they might attempt to enact their goals, by affecting the approach the terrorist group takes to causing existential harm. For example, the radical environmental group RISE sought to kill off humanity and repopulate the Earth with a small cadre of enlightened revolutionaries. Repopulation requires the Earth to still sustain life, so causing existential harm scenarios that leave the Earth inhospitable would not be desirable.

¹⁶ Gaia Liberation Front, *Statement of Purpose (A Modest Proposal)* <<http://www.churchofeuthanasia.org/resources/glf/glfsp.html>> (last accessed 16 February 2023).

¹⁷ Gula, “Eco-Kamikazes Wanted” (1989) IX(VIII) *Earth First! Journal* 21.

¹⁸ D Milbank, “A Strange Finnish Thinker Posits War, Famine as Ultimate ‘Goods’” (*Asian Wall Street Journal*, 1994), cited in P Torres, *Morality, Foresight, and Human Flourishing: An Introduction to Existential Risks* (Durham, NC, Pitchstone Publishing 2017).

¹⁹ See TJ Kaczynski, *Technological Slavery: The Collected Writings of Theodore J. Kaczynski, aka “The Unabomber”* (Port Townsend, WA, Feral House 2010).

²⁰ The authors thank Phil Torres for drawing their attention to this idea; see Torres, *supra*, note 18.

²¹ RN Smart (1958). “Negative Utilitarianism” (1958) 67(268) *Mind* 543.

²² Lifton, *supra*, note 2.

Nevertheless, motivation is not enough. A terrorist must have a pathway to cause existential harm.

III. Types of existential terrorism

Three main pathways exist for terrorists to cause existential harm: *existential attacks*, *existential spoilers* and *systemic harm*. In an existential attack, the terrorist aims to carry out an attack that causes existential harm either directly or indirectly through anticipated responses from target actors in a relatively short time period after the attack. An existential spoiler is a terrorist attack to intentionally sabotage, disrupt or otherwise prevent existential risk mitigation measures from succeeding. Systemic harms are terrorist attacks that may not be intended to cause existential harm but create global effects that render the international community unwilling or unable to conduct existential risk-reduction measures.

I. Existential attack

Small numbers of actors simply cannot by themselves launch an attack with humanity-ending consequences unless they employ a tremendous amount of asymmetric leverage. We assume this will involve advanced scientific, technological or social scientific awareness and understanding. The would-be world destroyers must either wield considerable technologies or be able to adroitly influence events such that others employ such technologies. Creating and dispersing a human-ending pathogen, unleashing an artificial superintelligence (ASI) or commandeering nuclear arsenals all require extremely high levels of technical and operational knowledge. A major factor in the long-term risk of existential attack is thus how the availability of technology and know-how evolves over time. Likewise, triggering a chain of events that starts a global war would require significant social scientific understanding about states' motivations and various levers by which to manipulate these.

a. Direct attack

i. Genetically engineered microorganisms. Of all the extant weapons that could pose a direct existential threat, the only ones that are even plausibly within the capability of terrorists to produce and deploy on their own are those based on biotechnology, mostly weapons consisting of pathogenic microorganisms. This is because, in the right environment, biological organisms are self-replicating and, in the case of contagious agents, self-propagating. This in turn implies that there is, at least at first glance, no upper limit to their spread, which makes them the ultimate asymmetric weapon for terrorists and other small groups or individual actors seeking to destroy the world.

There is little doubt that terrorists and other violent non-state actors brandishing biological weapons could bring about mass casualties and deaths, but could they do so on a scale that would be truly existential? Here the picture is not as clear, since there are several barriers – both natural²³ and resulting from our ability to respond²⁴ – that set a high bar for any disease that seeks to wipe out the human race. This is where concerns

²³ H McCallum, "Disease and the Dynamics of Extinction" (2012) 367 *Philosophical Transaction of the Royal Society B* 2828–39; R MacPhee and A Greenwood, "Infectious Disease, Endangerment, and Extinction" (2013) 2013 *International Journal of Evolutionary Biology* 571939.

²⁴ A Adalja, "Why Hasn't Disease Wiped Out the Human Race?" (*Atlantic*, 17 June 2016) <<https://www.theatlantic.com/health/archive/2016/06/infectious-diseases-extinction/487514/>> (last accessed 27 December 2022).

regarding biotechnology come in, since traditional microbiological techniques are unlikely, even with considerable effort, to produce a pathogen able to surmount these inherent obstacles. Modern biotechnology, however, especially what has come to be known as synthetic biology,²⁵ has witnessed several breakthroughs in the past three decades. These include the latest gene-editing technologies (especially CRISPR/Cas-9), rapid DNA sequencing²⁶ and DNA/RNA synthesis, which enable scientists to reliably and efficiently manipulate biological systems at the most basic levels.²⁷

Even more than the fundamental efficacy of these new tools, it is the deskilling that they engender that is central to their threat potential in the terrorist context. As these technologies advance, the levels of education, technical skill and experience required to engineer a microorganism are reduced.²⁸ For example, in 2017, a relatively small Canadian biotechnology company called Tonix synthesised the horsepox virus,²⁹ which is harmless to humans but is in the same family as smallpox, which decidedly is not. At the same time, globalising forces such as online education, marketplaces for used laboratory equipment and lowering costs of equipment mean that this deskilling can rapidly spread to even the most unstable regions of the world where terrorists and other violent groups thrive, thus further “democratising” the technology.³⁰ As an example, the International Genetically Engineered Machine (iGEM) competition brings together hundreds of teams of high school and undergraduate students annually from all over the world, and the winning entries represent levels of sophistication that several years ago were the exclusive domain of experienced personnel in advanced laboratories.³¹ Moreover, advances in biotechnology are inherently dual use, and, unlike nuclear weapons technology, they have multiple beneficial applications (especially in fighting disease) and employ techniques that quickly become ubiquitous.

It is feared that these factors, resulting from the enormous strides in modern biotechnology, will allow adversaries, even at the level of small non-state groups or perhaps individuals, to essentially first design a disease capable of existential levels of harm and then produce a pathogen with the required characteristics to fit the design.³² Such an engineered pathogen could conceivably disregard the generally observed inverse relationship between a pathogen’s lethality and its transmissibility.³³ Nor would aspiring world-killers be limited to traditional pathogens: novel harm agents, ranging from transmissible cancers and bacteriophages to genetic modifications of neurochemistry, could also be considered.

²⁵ Synthetic biology can be defined as the subset of biology that relates to the engineering of biological systems. See AA Cheng and TK Lu, “Synthetic biology: an emerging engineering discipline” (2012) 14 Annual Review of Biomedical Engineering 155–78.

²⁶ “Special Report, Synthetic Biology: Life 2.0” (*The Economist*, 2006) <<http://www.economist.com/node/7854314>> (last accessed 16 February 2023).

²⁷ Torres, *supra*, note 18, 50.

²⁸ JK Wickiser, KJ O’Donovan, M Washington, S Hummel and FJ Burpo, “Engineered Pathogens and Unnatural Biological Weapons: The Future Threat of Synthetic Biology” (2020) 13(8) CTC Sentinel.

²⁹ G Koblenz, “The De Novo Synthesis of Horsepox Virus: Implications for Biosecurity and Recommendations for Preventing the Reemergence of Smallpox” (2017) 15 Health Security 620–28.

³⁰ Torres, *supra*, note 18; V Madhwa, “The Genetic Engineering Genie Is Out of the Bottle” (*Foreign Policy*, 11 September 2020) <<https://foreignpolicy.com/2020/09/11/crispr-pandemic-gene-editing-virus/>> (last accessed 16 February 2023); R Zilinskas and P Mauger, “Occasional Paper 21: Biotechnology E-commerce: A Disruptive Challenge to Biological Arms Control” (*James Martin Center for Nonproliferation Studies*, 2015) <<https://nonproliferation.org/biotechnology-e-commerce-a-disruptive-challenge-to-biological-arms-control/>> (last accessed 16 February 2023).

³¹ Wickiser et al, *supra*, note 28; details of the 2022 iGEM competition can be found at <<https://jamboree.igem.org/2022>> (last accessed 27 December 2022).

³² R Casagrande, *Risk and Benefit Analysis of Gain of Function Research* (Takoma Park, MD, Gryphon Scientific 2015).

³³ Global Catastrophic Risks 2016 (2016 *Global Challenges Foundation*) <<http://globalprioritiesproject.org/wp-content/uploads/2016/04/Global-Catastrophic-Risk-Annual-Report-2016-FINAL.pdf>> (last accessed 16 February 2023).

The key question becomes how likely it is that terrorists or other misanthropic individuals will be able to realise biotechnology's most extreme potential for harm. Expert opinion differs. Some regard the threat as overblown, citing persistent technical obstacles to the realisation of the capabilities described above, especially for non-state actors.³⁴ After all, there is much we still do not know, especially in the realms of epigenetics, pleiotropic effects (one gene having multiple effects), polygenes (multiple genes acting in concert to have an effect) and epistatic (modifier) genes. Others contend that, "As the molecular engineering techniques of the synthetic biologist become more robust and widespread, the probability of encountering one or more of these threats is approaching certainty."³⁵ A recent study by the National Academies of Science, Medicine and Engineering described the bottlenecks to modifying and creating pathogens but acknowledged that most if not all of these could be overcome in time,³⁶ with most observers agreeing that the more complex and exotic achievements will take longer to arrive. However, even if all of the obstacles are not surmountable today, the rapid forward march of biotechnology suggests that, at some point in the not-too-distant future, terrorists might be capable of brandishing an existential-level microbe.

ii. Future technology weapons: artificial superintelligence and grey goo. While biotechnology is already widespread and advancing rapidly, there are also more speculative technologies that could conceivably be employed by terrorists as existential weapons. An ASI – an artificial intelligence capable of performing all the cognitive tasks a human is capable of better than a human can – could provide terrorists with the means to wreak existential harm. An ASI controlled or inspired by terrorists could conceivably use its superior intelligence to take over society, extract resources to make the Earth unliveable or exterminate humanity. Even a simple ASI designed to calculate the last digit of pi or create paperclips could cause existential harm because it may need "to acquire an unlimited amount of physical resources and, if possible, to eliminate potential threats to itself and to its goal system. Human beings might constitute potential threats; they certainly constitute physical resources."³⁷

Computer scientists are in strong disagreement over when and whether an ASI could be created.³⁸ Without knowing the technological pathways by which an ASI can be created, it is difficult to judge the general potential for terrorists to acquire and use an ASI. For example, if creating an ASI requires a massive data centre, advanced supercomputers or extraordinarily high programming skills, terrorist acquisition could be quite difficult, while if all that is required to create an ASI is a laptop and connection to the cloud, then this might make terrorist acquisition more likely. Even in the first case, if a radicalised computer scientist can easily build an ASI once given access to the data centre or supercomputer, then insider threats become quite significant. Although a growing body of research is exploring how to align artificial intelligence with humanity's well-being, a terrorist could simply ignore such guidance and develop an unaligned ASI without any safeguards.³⁹

³⁴ C Jefferson, F Lentzos and C Marris, "Synthetic Biology and Biosecurity: Challenging the Myths" (2014) 2 *Frontiers in Public Health* 115.

³⁵ Wickiser et al, *supra*, note 28.

³⁶ M Imperiale et al, *Biodefense in the Age of Synthetic Biology* (Washington, DC, National Academies Press 2018).

³⁷ N Bostrom, *Superintelligence* (Paris, Dunod 2017).

³⁸ S Armstrong and K Sotala, "How We're Predicting AI – or Failing to" in J Romportl, P Ircing, E Zackova, M Polak and R Schuster, *Beyond AI: Artificial Dreams* (Pilsen, University of West Bohemia 2012) pp 52–75; S Baum, A Barrett and RV Yampolskiy, "Modeling and Interpreting Expert Disagreement about Artificial Superintelligence" (2017) 41(7) *Informatica* 419–28.

³⁹ E Yudkowsky, "Creating Friendly AI 1.0: The Analysis and Design of Benevolent Goal Architectures" (*The Singularity Institute*, 15 June 2001) <<https://intelligence.org/files/CFAI.pdf>> (last accessed 14 June 2023); ME Castel, "The Road to Artificial Super-Intelligence: Has International Law a Role to Play?" (2016) 14(1) *Canadian Journal of Law and Technology*.

Self-replicating nanorobots created or released by terrorists represent another scenario for causing existential levels of harm. Nanotechnologist Eric Drexler theorised about self-replicating nanomachines that would replicate exponentially: “Imagine such a replicator floating in a bottle of chemicals, making copies of itself . . . in less than two days, they would outweigh the Earth; in another four hours, they would exceed the mass of the Sun and all the planets combined – if the bottle of chemicals hadn’t run dry long before.”⁴⁰ Of course, the nanobots do not need to consume the Earth, but only enough of the critical material that humanity needs to survive. But, in reality, nanobots are likely to have limits. The speed of replication, dispersion of the nanobots, energy requirements and material consumption will all affect the existential risk.⁴¹ An existential terrorist would require nanobots that can replicate very quickly, such that the probability of protective interventions, running out of energy or not acquiring enough material for replication is very low. To do this, the terrorist-released nanobot would likely need to be able to consume a broad range of materials (or be extremely broadly dispersed) so that localised overconsumption does not lead to agent starvation. In any event, terrorist acquisition of a nanobot capable of ecophagy seems likely to require extremely high technical knowledge and specialised equipment. Likewise, even if it is possible, future safety protocols may limit ecophagy risks.⁴²

Assessing the risk of a terrorist using a technology that does not yet exist is inherently fraught. Nonetheless, some observations can be made. First, any specialised resource requirements will place limitations on terrorists’ ability to acquire a future weapon, with the magnitude varying based on resource availability and quantity required. Second, the amount of technical knowledge required to use a future technology to manifest an existential risk will be a major factor in the risk; if the technology requires significant tacit knowledge generated from years of experience, the risk is much lower. Third, human-imposed limits such as safety standards are unlikely to be of much benefit in this context, since terrorists will likely choose to ignore them.

Finally, the risk will vary over time. An ASI or all-consuming nanobot has never been created, so the immediate risk is currently zero. It is possible that it will remain so. For example, some scientific researchers doubt whether creating molecular assemblers is even possible.⁴³ However, once the threshold for inventing these technologies has been crossed, the risk will immediately increase. As awareness and knowledge of the technology and how to reproduce it grows, the risk will grow too. At the same time, responsible – state or non-state – actors can be expected to take actions to reduce the risk, either through controls on the technology or counterterrorism to reduce the capacity of terrorists to exploit the technology. Of course, future technologies not yet envisioned may create additional pathways for exploitation that we cannot yet foresee.

In any event, terrorists do not necessarily require their own weapons.

b. Indirect attack

Terrorists might leverage the existing capabilities of states to bring about existential harm, in what we refer to as an *indirect* existential attack. This form of existential terrorism

⁴⁰ E Drexler, *Engines of Creation: The Coming Era of Nanotechnology* (New York, Anchor 1987).

⁴¹ RA Freitas, “Some Limits to Global Ecophagy by Biovorous Nanoreplicators, with Public Policy Recommendations” (Robert A. Freitas Jr, 2000) <<http://www.rfreitas.com/Nano/Ecophagy.htm>> (last accessed 14 June 2023).

⁴² O Häggstrom, *Here Be Dragons: Science, Technology and the Future of Humanity* (Oxford, Oxford University Press 2016).

⁴³ RE Smalley, “Of Chemistry, Love, and Nanobots” (*Scientific American*, 2001) <https://web.archive.org/web/20120723062135/http://www.sciamdigital.com/index.cfm?fa=Products.ViewIssuePreview&ARTICLEID_CHAR=F90C4210-C153-4B2F-83A1-28F2012B637> (last accessed 14 June 2023).

aims to trigger a sequence of actions that culminates in the manifestation of a direct existential threat. The most likely candidate in this regard is indirect nuclear terrorism.

i. Nuclear weapons. Even if terrorists eventually succeed – through theft or indigenous production – in acquiring their own nuclear weapons,⁴⁴ they will not plausibly be able to secure sufficient numbers to pose an existential threat. The remaining option is to exploit state arsenals and thus bring about a global nuclear war. There are two main pathways to catalyse nuclear war. The first, technical, path involves infiltrating nuclear weapon command-and-control systems and either launching multiple weapons or spoofing the system to ensure that one or more states launch their own arsenals. The second, political, path seeks to manipulate social and political systems to induce states to engage in large-scale nuclear exchanges. These are not mutually exclusive, and many scenarios along these lines involve a combination of both paths.

Scenarios of technical exploitation by terrorists of nuclear arsenals usually involve terrorists launching some sort of cyberattack that gives them access to secure nuclear systems. Although among the best-protected systems in the world, nuclear facilities in general are believed by many to be vulnerable to cyberattack, with so-called “air-gapping” from external networks not always maintainable in practice.⁴⁵ The increasing automation and networking of infrastructure generally are also likely to affect even the most secure systems, like those protecting nuclear weapons,⁴⁶ while the capabilities of malevolent hackers are continually increasing. The most extreme result of such digital intrusions would be for terrorists to gain control of operational nuclear weapons. There is extensive debate surrounding the security of so-called NC3 (nuclear command, control and communications) systems,⁴⁷ where, for instance, many experts argue that the US arsenal is relatively secure (perhaps because of its rather antiquated technology), but this is not necessarily the case for other countries with less robust controls, such as Pakistan.

Even if terrorists do not succeed in gaining enough control of systems to launch the nuclear weapons, there remains the possibility that they will gain sufficient access to spoof early-warning systems and thus provoke a catastrophic inter-state nuclear exchange. There have long been concerns with these early-warning systems (and several historic near-misses from false alarms), with particular concern focused on “Dead Hand” systems like those ostensibly used by Russia⁴⁸ and some concern given to China’s space-based early-warning system.⁴⁹ If a terrorist group or even individual could covertly trigger these systems by simulating, say, an impending missile strike⁵⁰ and convince a state that an enemy nuclear attack was in progress (especially if this is accompanied by additional forms

⁴⁴ G Allison, *Nuclear Terrorism: The Ultimate Preventable Catastrophe* (Ephrata, PA, Owl Books 2014); C McIntosh and I Storey, “Between Acquisition and Use: Assessing the Likelihood of Nuclear Terrorism” (2018) 62(2) *International Studies Quarterly* 289–300.

⁴⁵ CQ Choi, “Nuclear Cybersecurity Woefully Inadequate” (*IEEE Spectrum*, 7 October 2015) <<https://spectrum.ieee.org/nuclear-cybersecurity-woefully-inadequate>> (last accessed 14 June 2023); “The NTI Nuclear Security Index 5th Edition” (*Nuclear Threat Initiative*, 2020) <<https://www.ntiindex.org/>> (last accessed 14 June 2023).

⁴⁶ P Stoutland, “Growing Threat: Cyber and Nuclear Weapons Systems” (*Bulletin of the Atomic Scientists*, 18 October 2017) <<https://thebulletin.org/2017/10/growing-threat-cyber-and-nuclear-weapons-systems/>> (last accessed 14 June 2023).

⁴⁷ For an extensive discussion of NC3 issues, see the series of reports and interviews commissioned and hosted by the Institute for Security and Technology at <<https://securityandtechnology.org/ist-policy-lab/successes/nc3-systems-and-strategic-stability/>> (last accessed 16 February 2023). See also P Tucker, “Hacking into Future Nuclear Weapons: The US Military’s Next Worry” (*Defense One*, 29 December 2016) <<https://www.defenseone.com/technology/2016/12/hacking-future-nuclear-weapons-us-militarys-next-worry/134237/>> (last accessed 14 June 2023).

⁴⁸ A Barrett, *False Alarms, True Dangers?* (Santa Monica, CA, RAND Corporation 2016).

⁴⁹ E Heginbotham et al, *China’s Evolving Nuclear Deterrent: Major Drivers and Issues for the United States* (Santa Monica, CA, RAND Corporation 2017) p 143.

⁵⁰ Barrett, *supra*, note 48.

of disinformation), this could lead the state to erroneously launch its own first strike. Beyond the cyber dimension, even sending up physical objects like weather balloons might lead states with less sophisticated early-warning capabilities to initiate a counterstrike. Decisions to allow the autonomous launching of nuclear weapons using artificial intelligence would likely increase these risks due to the brittleness of the artificial intelligence.⁵¹

While existing controls such as permissive action links probably minimise the possibility of a single insider launching a nuclear weapon on their own, cyberthreats can be facilitated if the terrorist actor manages to obtain an insider within the nuclear weapons infrastructure. Insiders can bridge the air gap or open the proverbial digital doors to cyberintruders, or at the very least they can provide valuable information that assists a cyberattacker in circumventing defences. Great strides have been made in detecting insiders and preventing insiders from accessing nuclear weapons complexes in countries like the USA,⁵² but these systems are far from perfect, not to mention that other nuclear-weapon states might have less robust insider threat controls.

On the more political side, terrorists could manipulate a state into launching their nuclear arsenals against another state by convincing the military or political leadership that there is strategic advantage in this. The terrorist would likely need to mount a “false-flag” operation, where the terrorist group perpetrates an attack with large political and social consequences on a nuclear-weapon state and plants false evidence that a rival state was behind the attack. Recent technological developments that provide new tools for disinformation, ranging from “deep fakes”⁵³ to even intentionally triggering radiation detectors,⁵⁴ could facilitate such operations. Engaging in this type of behaviour in two or more states that are already both nuclear armed and engaged in a tense geopolitical rivalry would only heighten the risk. If a terrorist group were to get their hands on even a single nuclear weapon and use it in a false-flag attack, this would presumably enhance their odds of success, but even sizeable or symbolic conventional attacks in a suitably tense environment might be able to precipitate inter-state nuclear war.⁵⁵

These strategic machinations, while plausible, necessarily require certain actors (especially political or military elites) to respond in specific ways to particular stimuli. The vicissitudes of human behaviour suggest that the required chain of decisions and actions will not necessarily be linear, and therefore this avenue of existential terrorism becomes highly uncertain for the terrorist perpetrators. Furthermore, such operations are only likely to be successful in highly charged geopolitical contexts, which will only apply at certain points in time, thus making such a threat somewhat opportunistic in addition to unreliable.

ii. Redirecting planetary protection. Another plausible scenario is redirecting a benign planetary protection system. An asteroid one to two kilometres in diameter striking the Earth would generate an explosive force roughly equivalent to 1 million megatons of

⁵¹ Z Kallenborn, “Giving an AI Control of Nuclear Weapons: What Could Possibly Go Wrong?” (*Bulletin of the Atomic Scientists*, 1 February 2022) <<https://thebulletin.org/2022/02/giving-an-ai-control-of-nuclear-weapons-what-could-possibly-go-wrong/>> (last accessed 14 June 2023); J Johnson, “Catalytic Nuclear War’ in the Age of Artificial Intelligence & Autonomy” (2021) *Journal of Security Studies*.

⁵² M Bunn and S Sagan (eds), *Insider Threats* (Ithaca, NY, Cornell University Press 2017).

⁵³ A Zegart, “The Tools of Espionage Are Going Mainstream” (*The Atlantic*, 27 November 2017) <<https://www.theatlantic.com/international/archive/2017/11/deception-russia-election-meddling-technology-national-security/546644/>> (last accessed 14 June 2023).

⁵⁴ JH Leckey et al, *Special Nuclear Material Simulation Device*. U.S. Patent 8,804,898 (filed 12 August 2014).

⁵⁵ The authors are indebted to William Potter for this idea.

dynamite.⁵⁶ Such an impact would create a massive disruption to the ecosystem, destroy global agriculture and directly kill a substantial fraction of the world's population.⁵⁷ Thankfully, asteroids of this size or greater appear to be relatively scarce in our planetary neighbourhood, and their orbits are unlikely to bring them into serious risk of a collision with the Earth. If a planet-killing asteroid were identified, the National Air and Space Administration's (NASA) Double Asteroid Redirection Test (DART) mission and similar efforts are designed to push such an asteroid away from the Earth, perhaps through slight adjustments to the asteroid's orbit or simply by detonating a nuclear bomb on it.

However, any system intended to deflect asteroids could also conceivably be used to push them into Earth's orbit.⁵⁸ Terrorists could plausibly manipulate or commandeer planetary defence systems to deflect non-menacing near-Earth objects onto an Earth-impacting trajectory. This would admittedly be an incredibly difficult feat that would be highly contingent on various factors. For instance, it is likely that the terrorist would need to change target trajectories for a test of a planetary defence system without the changes being recognised or corrected (which is only relevant if planetary defence mission tests continue). In addition, the terrorist would need to ensure the payload for the mission is adequate to generate the force necessary to move the asteroid into a collision path. Both are likely to take extraordinarily high levels of technical knowledge of astrophysics and the internal processes of the organisation carrying out the planetary defence missions. In addition, the terrorist or terrorists would need to evade or prevent any safeguards aimed at confirming anticipated trajectories. The most plausible scenario is to have a well-placed insider sympathetic to an apocalyptic ideology or controlled by some means by an apocalyptic organisation. This scenario becomes more plausible to the degree that commercial space companies develop the capability to manipulate the trajectories of near-Earth objects, such as future space mining companies, assuming such manipulations generate adequate force to shift a near-Earth object into a collision orbit.⁵⁹

2. Existential spoilers

Existential terrorism can constitute something of a paradox: a terrorist can kill all of humanity without directly harming anyone. In a variant of the discussion above, imagine a planet-killer asteroid is identified and on a collision course with Earth. NASA prepares a planetary defence mission to shift the asteroid's orbit away from the Earth.⁶⁰ A terrorist group manages to place a sympathiser in NASA, who sabotages the mission. If the mission is delayed too long and the asteroid hits, then the terrorist group would have caused existential harm with no direct casualties.

Existential spoilers are thus terrorist attacks aimed at sabotaging, disrupting or otherwise preventing existential risk-mitigation efforts from succeeding. Alternatively, terrorist attacks may attempt to remove safeguards that protect against existential harms. Table 1 provides examples of existential spoilers for various types of existential risk.

⁵⁶ W Napier, "Hazards from Comets and Asteroids" in N Bostrom and MM Cirkovic (eds), *Global Catastrophic Risks* (Oxford, Oxford University Press 2008).

⁵⁷ C Sagan, "Dangers of Asteroid Deflection" (1994) 368 *Nature* 501.

⁵⁸ A Harris, G Canavan, C Sagan and S Ostro, "The Deflection Dilemma: Use vs. Misuse of Technologies for Avoiding Interplanetary Collision Hazards" (*NASA Jet Propulsion Laboratory*, 5 November 1993).

⁵⁹ M Petrova, "The First Crop of Space Mining Companies Didn't Work Out, but a New Generation Is Trying Again" (*CNBC*, 9 October 2022) <<https://www.cnbc.com/2022/10/09/space-mining-business-still-highly-speculative.html>> (last accessed 14 June 2023).

⁶⁰ J Handel and J Surowiec, "NASA Confirms DART Mission Impact Changed Asteroid's Motion in Space" (*NASA*, 11 October 2022) <<https://www.nasa.gov/press-release/nasa-confirms-dart-mission-impact-changed-asteroid-s-motion-in-space>> (last accessed 14 June 2023).

Table 1. Existential spoiler examples.

Existential risk	Existential spoiler example
Super-volcanoes	Disrupting volcano cooling measures
Comets and asteroids	Attacks on planetary defence missions or tests
Cosmic explosions	Attacks on astro-engineering constructs or sanctuaries
Climate change	Attacks to disrupt climate change talks or geo-engineering efforts
Pandemics	Attacks on vaccine researchers, manufacturers or distributors during a lethal pandemic
Artificial intelligence	Attempts to remove safeguards around artificial general intelligences or superintelligences through sabotage or violence
Social collapse	Attacks aimed to impede social resilience
Nuclear war	Attempts to disrupt peace talks between warring nuclear powers
Nanotechnology	Attacks on researchers developing mitigation measures to nanorobotics

The motivations underlying engaging in an existential spoiler event do not always need to be existential themselves. Most terrorists are unlikely to seek the destruction of humanity: terrorists seeking narrow changes to their home country's policies around abortion would have no reason to spoil risk-mitigation measures. However, it is plausible that sometimes a terrorist's ideology might align with seeking to disrupt existential risk-mitigation efforts, even though the terrorist does not possess motivations to inflict existential harm per se. For example, radical environmentalists concerned about the harm that geo-engineering does to the environment may seek to sabotage or disrupt those efforts without any desire to cause or prevent existential harm. Of course, this alignment is likely to be highly idiosyncratic and require consideration of how specific terrorist ideologies interact with specific existential risks.

For those instances where there is little to no alignment between the terrorist's ideology and the object of the risk-mitigation measure, spoiler attempts will likely require high existential motivation (eg terrorists are not typically concerned about near-Earth objects, so spoiler actions in this instance would likely be directed intentionally towards causing existential harm). The motivation may be in either direction: causing or preventing the existential harm from manifesting. For example, an Aum Shinrikyo-like organisation may desire the existential catastrophe to occur and so sabotage the mitigation efforts. Conversely, a radical environmental group may sabotage climate change efforts because they fear global consolidation around suboptimal (or harmful) risk-reduction measures.

Executing a spoiler attack is unlikely to require significant capabilities, although the exact details will depend on the target. Disrupting a major international summit would likely require little more than explosives or small arms. Of course, such a summit can be expected to have higher levels of security, which a terrorist would need to overcome, but such a problem is hardly insurmountable. The challenge is attacking at a point in time and space where the summit could not easily be rescheduled. Some specific forms of attack such as removing safeguards around an ASI or sabotaging geo-engineering attempts will require a stronger technical understanding regarding how those systems work and potential vulnerabilities. But it is not obvious that the technical capabilities required to

carry out the sabotage would be unusual to find within many terrorist organisations, or even among lone actors or small cells.

Whether the terrorist actually succeeds in causing existential harm from their spoiler attempts will be highly contingent. An existential risk-mitigation measure is most significant when an existential risk is on the verge of manifesting or the risk-mitigation effort cannot just be rescheduled or modified. If the mitigation measure can be fixed or alternatives developed before the risk manifests, then the only harm is the resource costs and any direct victims of the attack. For example, extreme climate change is an existential risk that manifests over the course of decades or centuries. If terrorists disrupt a major climate change summit, global leaders could just hold another one a few weeks or even years later. Conversely, if a strike from a near-Earth object of sufficient size is imminent, delaying a planetary defence mission by a few days may result in existential harm. Thankfully, planet-killer asteroids that could hit Earth are rare: astronomers in 2022 detected an asteroid, 2022 AP7, that is large enough to destroy humanity (2022 AP7 is 1.1–2.3 km in diameter), but they do not believe the asteroid will cross Earth’s orbit for centuries.⁶¹ So, disrupting a planetary defence mission today might have no effect, while disrupting a mission in the year 2300 might be existential. Knowing when and how to take advantage of such windows of opportunity may itself require at least some technical and scientific knowledge to anticipate them.

3. Systemic harm

Unlike the three previous types of existential terrorism, the last type – systemic harm – does not arise from a conscious goal of the terrorists to bring about global destruction or hinder risk-mitigation efforts. Here, other terrorist actions have the unintended side effect of increasing the overall level of existential threat. There are three flavours that such systemic harm might come in, although these are not mutually exclusive:

- *Impeding mitigation*: This is similar to the spoiler category above in that it has the same overall effect but is not a purposeful consequence of the terrorists’ actions. In most cases this will involve the curtailment of scientific or policy advances that would prevent or mitigate other natural or human-made existential risks. An example would be if a terrorist group used synthetic biology to create a pathogen which it used to perpetrate a mass-casualty (but not existential) attack. As a result of this attack and security concerns regarding similar future attacks by terrorists, one or more governments impose moratoria or extremely onerous controls on further synthetic biology research. This in turn could lead to scientists not having the tools or knowledge to adequately deal with a new, extremely virulent naturally occurring global pandemic or a military use of bioweapons that runs amok. The original terrorist action, which was intended to cause limited harm, would thus end up preventing the development of the capability to deal with the actual existential threat.
- *Opportunity cost*: This is a less direct version of systemic harm. In this case, one or more (major) terrorist attacks lead one or several of the world’s major powers to fixate policy attention on and devote enormous resources to countering the real or perceived threat of terrorism. This acts as a distraction from dealing with

⁶¹ N Davis, “Huge ‘Planet Killer’ Asteroid Discovered – and It’s Heading Our Way” (*The Guardian*, 1 November 2022) <<https://www.theguardian.com/science/2022/nov/01/huge-planet-killer-asteroid-discovered-and-its-heading-our-way>> (last accessed 14 June 2023); SS Sheppard et al, “A Deep and Wide Twilight Survey for Asteroids Interior to Earth and Venus” (2022) 164(4) *The Astronomical Journal* 168.

various existential threats. For instance, if the USA (similar to after 9/11) reacts to a large terrorist attack on its soil by steering the national security ship of state once again to focus on counterterrorism, then both legislative attention and budgets could be less available to prepare for, say, planetary defence against near-Earth objects. If this ultimately means that the world is less prepared to deal with an approaching asteroid or comet, the terrorist attack will have indirectly exacerbated the existential risk.

- *Overreaction*: State response to a major terrorist attack could itself increase existential risk. For instance, if a state decides that it should deploy an ASI to hunt down a terrorist group and hastily removes safeguards on the ASI's capabilities, the ASI itself might turn into an existential threat. Alternatively, longer, drawn-out terrorist campaigns that result in a vicious cycle of terrorist attack followed by government repression leading to further terrorist attacks might destabilise society. If this cycle is sufficiently prolonged and occurs in a sufficiently large number of countries simultaneously, it could at the extreme weaken global society and make other existential risks more likely to manifest or remain unmitigated.

IV. Conclusion

Existential terrorism represents the most extreme form of terrorism imaginable: terrorists who seek to destroy humanity. Devoting serious thought to the possibility of terrorists destroying the world may at first seem to be a somewhat frivolous and wasteful exercise. Yet the immeasurable consequences of any existential risk justify at least a preliminary exploration of threat feasibility, even if we ultimately conclude that the threat is negligible. Furthermore, as we have seen, there are often facets and complexities associated with the topic that only reveal themselves after one attempts an honest examination.

However, assessing the risk of existential terrorism is a challenging task. Like other forms of existential risk, direct empirical evidence for existential terrorism is impossible, by definition. If a terrorist actor succeeds in killing off humanity, then no human will be alive to study it. Existential terrorism thus presents an extreme form of the Irish Republican Army's maxim that "You have to be lucky all the time. We only have to be lucky once."⁶² At best, risk assessment must rely on crude proxies, such as understanding potential vulnerabilities that could be exploited and cases of failed existential terrorism attempts. Even though the failed attempts may be inherently biased (since a successful attempt may not possess the attributes of the failures), past cases can at least illustrate the pitfalls that a would-be existential terrorist must avoid.

While a detailed assessment of the risk of existential terrorism remains a task for future research, a preliminary examination of the relative risk of each of the above pathways can be made using a modified form of the classic risk formulation that combines measures of likelihood (modelled as the combination of motivation and capability) with vulnerability.⁶³ We can thus look at:

⁶² L Clutterbuck, "Terrorists Have to Be Lucky Once; Targets, Every Time" (*RAND Corporation*, 30 November 2008) <<https://www.rand.org/blog/2008/11/terrorists-have-to-be-lucky-once-targets-every-time.html>> (last accessed 14 June 2023).

⁶³ Consequences are omitted because existential risk has maximal consequences by definition.

Table 2. Preliminary assessment of the relative risk of existential terrorism pathways.

	Existential motivation requirements	Capability requirements	Contingency dependence
Existential attack (direct)	High	Extremely high	Low
Existential attack (indirect)	High	High	Medium to high
Existential spoiler	Low to high	Low to medium	Very high
Systemic harm	Low	Low to high	Low to very high

Note: The broad ranges for capability requirements and contingency dependence are because the two factors are likely to be inversely linked. That is, a terrorist might create systemic harm with low capabilities, such as a crude assassination that ignites a great power war, but generating systemic harm is likely to be very contingent. Conversely, attacks that are less contingency dependent are likely to require much higher capabilities (eg detonating a nuclear weapon).

- *Existential motivation required*: The degree to which the terrorist organisation is specifically motivated to cause existential harm;
- *Capability requirements*: The resources, technical knowledge and other organisational capabilities needed to enable the terrorist to carry out the attack; and
- *Contingency dependence*: The degree to which an existential terrorist attack is dependent on specific temporal, environmental, political, spatial or other contingencies.

Table 2 provides an initial assessment of the approximate levels of each of these risk dimensions that is likely to be required for each existential terrorism pathway. These values are informed by the above discussions of each pathway, although more rigorous future analysis is needed to confirm these. Moreover, these assessments, in particular the capability requirements, may change over time, especially if states adopt protective measures to reduce the risk.

Overall, we conclude that several plausible pathways exist for terrorists to destroy human civilisation, although the likelihood at present of any of them is very low. Within the bounds of feasibility (sometimes barely so), terrorists could conceivably develop genetically engineered microbes, catalyse nuclear war or, in the future, utilise novel technologies like ASIs and nanorobotics to carry out existential attacks. However, in the near to medium term, this is likely to require significant amounts of technical and scientific expertise and resources, far beyond a typical (or even state-sponsored) terrorist organisation. Of course, future technological advances – artificial intelligence and rapid prototyping are noteworthy examples – and other factors may lower the barriers considerably. Alternatively, and far more concerning, is the potential for terrorists to spoil existential risk-mitigation measures, such as disrupting planetary defence missions. However, the effectiveness of such attempts would be dependent on an impending existential harm manifesting through other means and is thus highly contingent on extraneous conditions. The contingency dependence is again high for causing systemic harm that undermines the ability to deal with other existential risks. Like high capability thresholds, high contingency thresholds also imply lower likelihood overall.

Yet, given the outsized consequences, even extremely low-probability existential terrorism should be at least considered by governments in their long-term policy. First, as governments consider how best to monitor and reduce the risk, they need to understand how the risk dynamics vary across each pathway to existential harm. Second, governments should monitor the availability of and access to technology and expertise that would

Box I. Global totalitarian government and existential terrorism

Within the existential risk literature, there are some advocates for including a stable totalitarian world order as an existential risk, for although humanity would continue physically, such a system would prevent humankind from reaching its full potential as a species.⁶⁴ While defining existential risk in this way is controversial, it is worthwhile considering the implications for the concept of existential terrorism.

First, it would undoubtedly expand the set of potential perpetrators. History has offered up no shortage of malignant narcissists or megalomaniacs who seek absolute power and to dominate and control others on a mass scale. It is overwhelmingly likely that, were they to acquire the means and opportunity, personages like Hitler or Stalin would have sought to bring the entire globe under totalitarian control. There is no reason to assume that future non-state would-be leaders in this mould (such as the leader of a far-right or far-left terrorist group or an erratic tech billionaire) would not seek to achieve Bondian global overlord status if they believed it was possible or the opportunity presented itself.

Second, it has obvious implications for the type of existential attack to expect. Given that imposing totalitarian control over the world does not require humanity to be killed off, a perpetrator may seek a cataclysmic war that enables the chosen elite to seize global control or release an artificial superintelligence that the actor believes is aligned to its ideological preferences but would not act to prevent an asteroid collision that would destroy the Earth in its entirety.

Last, and perhaps most significantly, the overreaction type of systemic harm looms large in the totalitarian dystopia form of existential risk. Indeed, one or more terrorist attacks could serve as the catalyst for an entrenched totalitarianism that in itself would constitute an existential risk. Scenarios under this category generally involve some combination of two dynamics. In the first, in the wake of a major terrorist attack or series of attacks and out of a genuine concern for security, governments adopt technologies that allow for ubiquitous surveillance. This “Big Brother” scenario, which facilitates totalitarian control, is then exploited by authoritarians and subsequently degenerates into totalitarianism. The other scenario involves opportunistic totalitarian actors utilising a terrorist attack directly as a justification to erode freedoms and civil liberties. The Reichstag Fire that allowed the Nazis to assume exclusive control of Germany is the canonical example in this regard.⁶⁵

enable existential harm, such as synthetic biology and artificial intelligence, although some of this is already done in the context of ordinary counterterrorism activities. Third, existential spoiler attacks are potentially easier for terrorist groups to pull off, but they are also highly contingent, and few terrorists will be directly motivated to make such attempts. Governments should thus pay close attention to this possibility when circumstances arise in which the actions of spoilers could be catastrophic. Systemic harms are also worthy of attention, although prevention can be expected to look much like normal counterterrorism. Nonetheless, some policy approaches will be useful across all three pathways.

1. Intelligence and information sharing

States should create a dedicated existential terrorism awareness function within their intelligence and/or law enforcement agencies to prioritise understanding and identifying terror groups with existential motivations. These can take various forms, from separate units to merely a separate set of responsibilities. This function requires monitoring at least

⁶⁴ B Caplan, “The totalitarian threat” in N Bostrom and MM Cirkovic (eds), *Global Catastrophic Risks* (Oxford, Oxford University Press 2008).

⁶⁵ L Boissoneault, “The True Story of the Reichstag Fire and the Nazi Rise to Power” (*Smithsonian Magazine*, 21 February 2017) <<https://www.smithsonianmag.com/history/true-story-reichstag-fire-and-nazis-rise-power-180962240/>> (last accessed 14 June 2023).

three types of organisation: known terrorist groups with explicitly existential motivations; existential-orientated factions within known terrorist groups minimally concerned with existential issues; and organisations with active apocalyptic beliefs that have not engaged in terrorist activity. The third type is likely to create difficult counterterrorism challenges regarding not violating important free exercise of religion and speech rights in open societies.

These dedicated functional apparatuses may also liaise with other global governments and international organisations to share information about terrorists within their states and the activities of specific threat actors. In practice, the unit performing this function will probably not be limited to the existential terrorism awareness task, given the rarity of the threat, but it may also provide subject matter expertise with respect to actors with existential motivations, such as apocalyptic cults. For example, Aum Shinrikyo's attack with the sarin nerve agent on the Tokyo subway system is significant, even if it did not cause existential harm. Alternatively, existential terrorism could be a specialisation available to any intelligence or law enforcement officers interested in mass-casualty terrorism. Such specialists may be brought together on an ad hoc basis to consider particular threats. This would reduce the organisational burden while increasing the onus on leaders to judge when counter-existential terrorist efforts should be formalised or combined.

2. Building resilience and protection

States also need to protect risk-reduction measures from potential terrorist attack. This requires understanding a measure's vulnerability to disruption and the criticality of the measure to broader risk reduction. That is, a risk-reduction measure that is unique and needs to take place at a particular time or place to be successful is far more vulnerable than a measure that can be easily delayed, requires general collective action or has numerous alternative measures. In practice, protecting risk-mitigation measures may look much like ordinary counterterrorism efforts – protecting a major climate change or ASI summit likely needs the same type of security as any major global summit between world leaders – but may require increasing event or location security at the most critical times and taking into account in its threat assessment the specific potential adversary objective of disrupting the mitigation measure. Building strong security cultures in entities that do not traditionally focus on security threats may also be necessary: for example, although NASA does do some classified work, as a primarily civilian agency it is unlikely to have the same emphasis on security as a military or intelligence agency. Although the scientific openness of NASA as a whole should not change, serious attention should be paid to assessing and plugging any potential security vulnerabilities related to existential risk-reduction efforts, such as planetary defence missions. This could entail such actions as classifying critical details about future missions. Some unique measures may be needed, such as identifying insider threats exhibiting extreme misanthropy. Similarly, security personnel need to assess the consequences and risks of counterterror responses to the risk-mitigation effort, since delaying a mitigation effort may result in existential harm. For extremely critical existential risk-reduction measures, states should focus on resilience and redundancy. This could include such coordinated actions as the USA encouraging allied states to conduct their own planetary defence tests or constructing additional seed banks to reduce risks of single points of failure.

3. A moral obligation

Existential terrorist groups do not threaten just one state, people, religion or corporation; they threaten all of humanity. States, local governments, corporations and communities

have – and the international community should recognise – a moral obligation to combat terrorists who desire existential harm. States have an obligation to conduct appropriate counterterrorism operations against such groups or individuals, including investigating, disrupting, arresting and, if appropriate, otherwise neutralising members. Conversely, the moral obligation also implies an obligation for members of the international community to provide financial, technological, intelligence and other aid to one another in the context of existential terrorist threats. That should include even adversarial states, albeit this will probably be limited in practice to intelligence sharing regarding specific threats.

The moral obligation to combat existential terrorism also raises larger moral questions about what other values states may be willing to or should be obligated to sacrifice. The USA justified a wide range of morally questionable actions like torture, violations of civil liberties and the invasions of Iraq and Afghanistan to prevent the harms from a future September 11 terrorist attack. If a state believed the very survival of humanity were at risk, what actions might this belief justify? Mass torture? Genocide? Nuclear war? What if a state does not believe humanity is at risk but uses existential terrorism concerns to justify brutality? These questions are far too large to explore here, but they require serious thinking, especially on how to reduce risks of abuse using existential terrorism as a smokescreen.

Although most versions of existential terrorism are extraordinarily unlikely, scenarios exist in which motivated terrorists leveraging modest capabilities could cause existential harm. There are broader circumstances in which terrorists or other violent non-state actors could either wittingly or unwittingly otherwise act to exacerbate existential risk. Policymakers should prioritise identifying and preventing the most likely among these. Those scenarios are the ones worth worry.

Acknowledgments. The authors would like to thank Seth Baum, Philipp C. Bleek, Nicholas Colosimo, Gregory Koblenz, Herbert Tinsley and attendees of Maastricht University’s workshop on long-term risks and future generations for their feedback on the paper. Any remaining errors or poor turns of phrase are the authors’ own. The views expressed in the article reflect only those of the authors and not any current or former employers, funders or affiliates.

Competing interests. The authors declare none.