GLOBAL PRIORITIES PROJECT 2017

Existential Risk

Diplomacy and Governance

GLOBAL PRIORITIES PROJECT

Future of Humanity Institute
UNIVERSITY OF OXFORD

FORMIN.FINLAND.FI
MINISTRY FOR FOREIGN AFFAIRS OF FINLAND

Table of contents

Authors

Sebastian Farquhar
John Halstead
Owen Cotton-Barratt
Stefan Schubert
Haydn Belfield
Andrew Snyder-Beattie

Acknowledgements

This report has benefited from the input of many minds. We would like to especially thank for their comments and suggestions Dr. Stuart Armstrong, Dr. Seth Baum, Andrei Botez, Dr. Niel Bowerman, Dr. Genya Dana, Carrick Flynn, Ben Garfinkel, Professor Timo Goeschl, Professor Lawrence Gostin, Dr. Petri Hakkarainen, Dr. Alan W. Harris, Professor Alan Harris, Dr. Hauke Hildebrandt, Dr. Hiski Haukkala, Professor David Heymann, Professor Anna-Maria Hubert, Antti Kaski, David Kelly, Professor David Keith, Dr. Raija Koivisto, Dr. Tim Kruger, Dr. Tom Inglesby, Professor Marc Lipsitch, Professor Mikhail Medvedev, Professor Adrian Melott, Dr. Piers Millett, Professor Juan Moreno-Cruz, Luke Muelhauser, Dr. Sean O'Heigeartaigh, Dr. Toby Ord, Dr. Andy Parker, Professor Edward Parsons, Professor Raymond Pierrehumbert, Dr. Ossi Piironen, Professor Steve Rayner, Dr. Sinikukka Saari, Carl Schulman, Dr. Pia-Johanna Schweizer, Dr. Jesse Reynolds, Dr. Norbert Reez, Dr. Catherine Rhodes, Professor Alan Robock, Professor Alan Ross, Hannah Sehan, Stephan de Spiegeleire, Jaan Tallinn, Dr. Theo Talbot, Professor Brian C. Thomas, Professor Brian Toon, Kevin Wong and the students and staff of the Geneva Centre for Security Policy.

In addition, we are grateful to the Ministry for Foreign Affairs of Finland who provided the funding which made this project possible and whose support and advice improved the outcome immeasurably.

Executive summary

The 2015 Paris Agreement represented a huge global effort to safeguard future generations from damaging climate change. But climate change is not the only serious risk to humanity. Our collective commitment to our children and future generations needs to extend to all existential risks — those with the potential to permanently curtail humanity's opportunity to flourish. These risks include nuclear war, engineered pandemics, and other catastrophes resulting from emerging technologies.

These disasters could cause an almost unimaginable loss. They would lead to immediate harm, but in their most extreme forms, they have the potential to wipe out humanity entirely.

Such risks may seem unlikely and distant. Indeed, in any one year they are improbable. But small probabilities accumulate - and because disaster risk reduction is a global public good individual nations will tend to underinvest in it. Nuclear weapons and climate change themselves would have once been unimaginable. It may be that emerging technologies introduce new risks that are even harder to manage. Managing existential risk may prove to be the decisive geopolitical challenge of the 21st century.

The first half of this report offers an overview of existential risks. The second half presents three opportunities for humanity to reduce these risks. These were chosen with the help of over 50 researchers and policy-makers out of more than 100 proposals emerged from three workshops at the University of Oxford and the Ministry of Foreign Affairs in Helsinki.

For each of these opportunities, humanity will require increasing levels of trust and international collaboration in order to face the challenges that threaten us all. Moreover, these risks are constantly evolving, and understanding them will need deep and sustained engagement with the global research community.

We hope that this report will go some way to advancing the discussion about the management of existential risks, and inspire action from well-placed individuals and institutions.

DEVELOP GOVERNANCE OF GEOENGINEERING RESEARCH

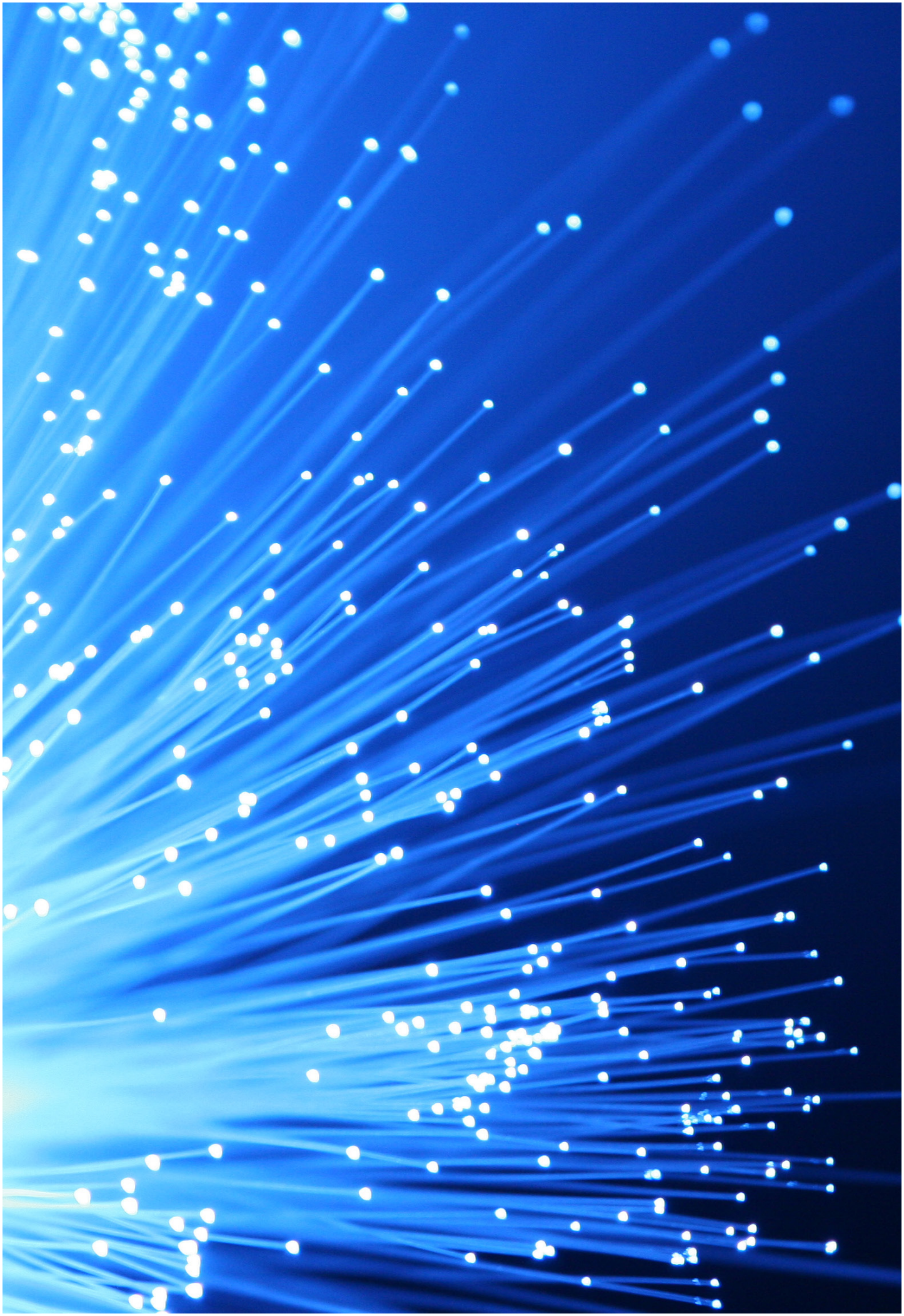
Geoengineering technologies like Solar Radiation Management have the potential to mitigate risks from climate change, while at the same time posing risks of their own. The current lack of international norms on acceptable research practices may well be holding back safe exploration of climate engineering options.

ESTABLISH SCENARIO PLANS AND EXERCISES FOR SEVERE ENGINEERED PANDEMICS AT THE INTERNATIONAL LEVEL

Existing scenario planning focuses on modest outbreaks at a mostly national level. As the 2015 Ebola outbreak showed, nations do not respond in isolation. Planning must become increasingly international, and should prepare for low-probability high-impact scenarios of pathogens synthesised to be more harmful than any naturally occurring disease.

BUILD INTERNATIONAL ATTENTION AND SUPPORT FOR EXISTENTIAL RISK REDUCTION

Existential risks are typically transnational and intergenerational. Overcoming them will need creative solutions to collective action problems, and shared political will. This will require the international community to build international capacity and draw the attention of national governments and international organisations to existential risk.



1. An introduction to existential risks

In day-to-day life, we all navigate a range of risks: each time we cross the road, for example, we face a relatively slight chance of serious injury or death. Some risks are more serious than others, so we devote more of our time and effort to mitigating them. For instance, other things being equal, it makes sense to devote more effort to reducing the risks of dying in a car accident than to avoiding contracting a rare and relatively harmless illness.

The seriousness of a risk depends on three things: its *scope* (the number of people it would affect), *severity* (how badly these people would be affected), and *probability* (how likely it is to occur). The diagram on page 7 gives examples of risks categorised according to scope and severity.[1] Policymakers in the international community work across each of these types of risk.

In this report, we focus exclusively on existential risks - those with the widest possible scope and the greatest possible severity, which are represented by the top right corner of the diagram. An *existential risk* is a risk that threatens the premature extinction of humanity or the permanent and drastic destruction of its potential for desirable future development.[2] Note that, on this definition, an existential risk need not actually kill everyone. For example, if a global catastrophe leaves some survivors alive but unable to rebuild society, then it would still qualify as an existential catastrophe.

Existential risks are especially worth focusing on because of their impact on the long-term future of humanity. For individuals, premature death is concerning because it would deprive them of a future which would otherwise last for many decades. In a similar way, premature extinction matters because it would deprive humanity of a future potentially lasting a million years or more. The sheer scale of the future at stake makes reducing existential risk hugely valuable.

1.1. AN OVERVIEW OF LEADING EXISTENTIAL RISKS

Over the course of the 200,000 year history of our species, humanity has been at risk of extinction as a result of natural catastrophes, such as asteroids and super-volcanic eruptions. *Anthropogenic* – human-caused – risks are a much newer phenomenon. Technological progress can give us the tools to improve society and to reduce existential risk, for example by providing the means to deflect large asteroids. However, technologies can also create new risks: with the invention of nuclear weapons, humanity gained the practical capacity to bring about its own extinction for the first time. A crucial political task for the international community will be to manage technological progress so that we enjoy the benefits while minimising the risks of existential catastrophe.[3]

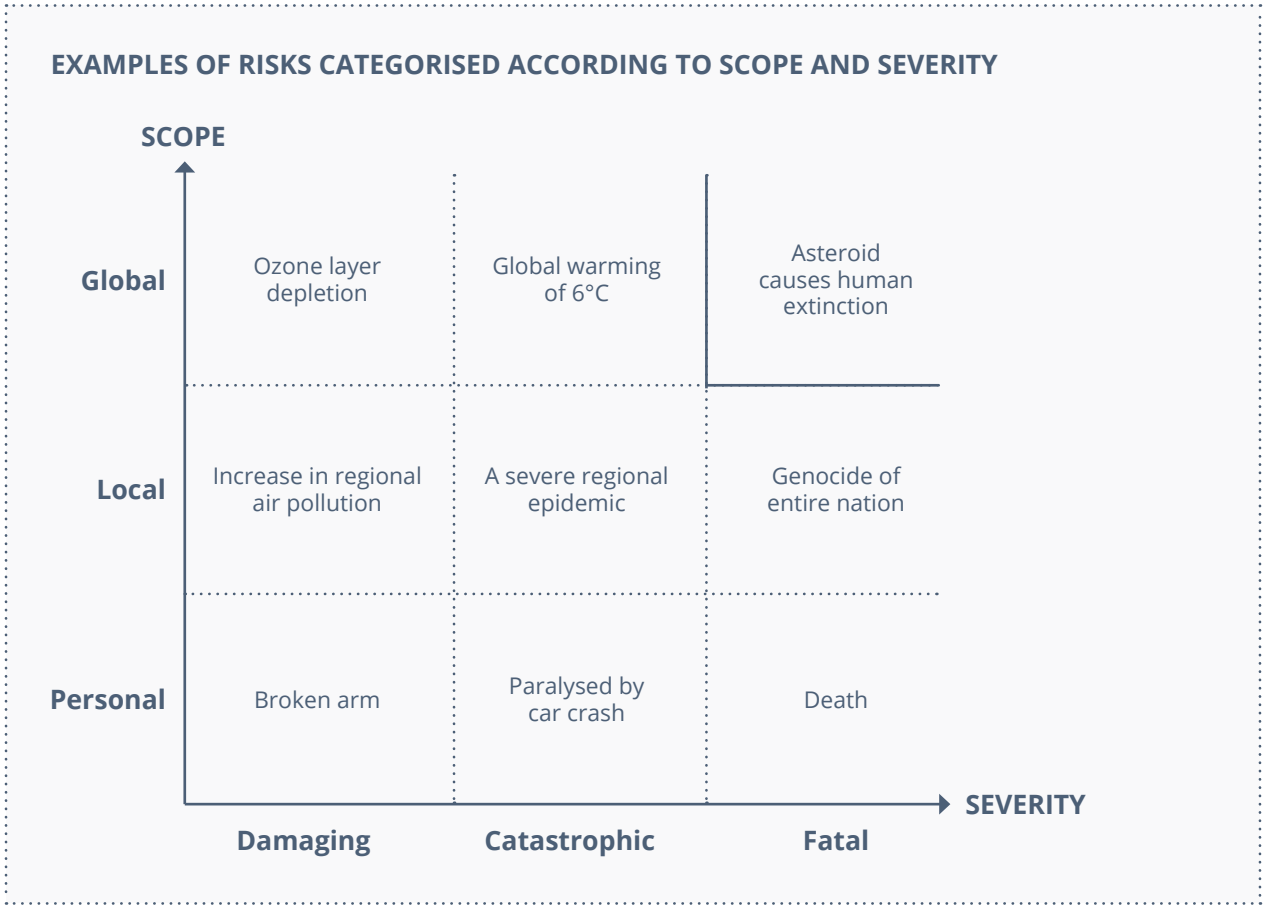
This also highlights the importance of focusing the attention of research communities on existential risk. Because many of these risks are from emerging technologies, humanity should not necessarily expect to already have the knowledge and tools needed to manage them. As a result, research into existential risks needs to be an early priority.

Existential risk is a deeply complex category. From asteroid strikes to extreme climate change to engineered viruses, they are not limited to one scientific domain. Some are instantaneous, whereas others play out over long periods of time. Some risks could only either kill everyone or be diverted altogether, whereas others might ultimately fall between these two extremes, causing a non-existential global or localised catastrophe. And though some, such as asteroid impacts, may be diverted by a single powerful actor, others, such as nuclear war, will require cooperation by most or all of the world's nations.

Not all risks are equally probable, and we can get some ideas of which are most likely. The historical record of natural risks gives us some ways to estimate how likely they are, and suggests that it is very unlikely that such events will extinguish humanity in the next 100 years.[4] Instead, it may be anthropogenic risks that pose the greatest threat.

In the remainder of this section, we provide a brief overview of the most serious existential risks, though we do not intend this list to be exhaustive.[5] We discuss risks which are directly existential, but also possible events which, while catastrophic, are not immediately existential. These are included because less severe catastrophes could bring about extinction indirectly and over a longer timeframe, for a number of possible reasons:

- Since there are no precedents in the last few thousand years of civilisation for events leading to the death of more than 20% of the global population,[6] it is very difficult to know how civilisation would respond to such a catastrophe.[7]



- There may be interaction effects between the risks which could cause one to cascade into a series of connected risks.
- Catastrophes that would be possible to recover from might make society less resilient to other risks.

On the other hand, there might be strong forces which make it likely that society will recover from all but the most severe catastrophes. We remain neutral on this debate and accordingly discuss the most severe, as well as somewhat less severe, catastrophes.

1.1.1 Nuclear war

The bombings of Hiroshima and Nagasaki demonstrated the unprecedented destructive power of nuclear weapons. However, even in an all-out nuclear war between the United States and Russia, despite horrific casualties, neither country's population is likely to be completely destroyed by the direct effects of the blast, fire, and radiation.[8] The aftermath could be much worse: the burning of flammable materials could send massive amounts of smoke into the atmosphere, which would absorb sunlight and cause sustained global cooling, severe ozone loss, and agricultural disruption – *a nuclear winter*.

According to one model [9], an all-out exchange of 4,000 weapons[10] could lead to a drop in global temperatures of around 8°C, making it impossible to grow food for 4 to 5 years. This could leave some survivors in parts of Australia and New Zealand, but they would be in a very precarious situation and the threat of extinction from other sources would be great. An exchange on this scale is only possible between the US and Russia who have more than 90% of the world's nuclear weapons, with stockpiles of around 4,500 warheads each, although many are not operationally deployed.[11] Some models suggest that even a small regional nuclear war involving 100 nuclear weapons would produce a nuclear winter serious enough to put two billion people at risk of starvation,[12] though this estimate might be pessimistic.[13] Wars on this scale are unlikely to lead to outright human extinction, but this does suggest that conflicts which are around an order of magnitude larger may be likely to threaten civilisation. It should be emphasised that there is very large uncertainty about the effects of a large nuclear war on global climate. This remains an area where increased academic research work, including more detailed climate modelling and a better understanding of how survivors might be able to cope and adapt, would have high returns.

It is very difficult to precisely estimate the probability of existential risk from nuclear war over the

next century, and existing attempts leave very large confidence intervals. According to many experts, the most likely nuclear war at present is between India and Pakistan.[14] However, given the relatively modest size of their arsenals, the risk of human extinction is plausibly greater from a conflict between the United States and Russia. Tensions between these countries have increased in recent years and it seems unreasonable to rule out the possibility of them rising further in the future.

1.1.2 Extreme climate change and geoengineering

The most likely levels of global warming are very unlikely to cause human extinction.[15] The existential risks of climate change instead stem from *tail risk* climate change – the low probability of extreme levels of warming – and interaction with other sources of risk. It is impossible to say with confidence at what point global warming would become severe enough to pose an existential threat. Research has suggested that warming of 11-12°C would render most of the planet uninhabitable,[16] and would completely devastate agriculture.[17] This would pose an extreme threat to human civilisation as we know it.[18] Warming of around 7°C or more could potentially produce conflict and instability on such a scale that the indirect effects could be an existential risk, although it is extremely uncertain how likely such scenarios are.[19] Moreover, the timescales over which such changes might happen could mean that humanity is able to adapt enough to avoid extinction in even very extreme scenarios.

The probability of these levels of warming depends on eventual greenhouse gas concentrations. According to some experts, unless strong action is taken soon by major emitters, it is likely that we will pursue a medium-high emissions pathway.[20] If we do, the chance of extreme warming is highly uncertain but appears non-negligible. Current concentrations of greenhouse gases are higher than they have been for hundreds of thousands of years,[21] which means that there are significant unknown unknowns about how the climate system will respond. Particularly concerning is the risk of positive feedback loops, such as the release of vast amounts of methane from melting of the arctic permafrost, which would cause rapid and disastrous warming.[22] The economists Gernot Wagner and Martin Weitzman have used IPCC figures (which do not include modelling of feedback loops such as those from melting permafrost) to estimate that if we continue to pursue a medium-high emissions pathway, the probability of eventual warming of 6°C is around 10%,[23] and of 10°C is around 3%.[24] These estimates are of course highly uncertain.

It is likely that the world will take action against climate change once it begins to impose large costs on human society, long before there is warming of 10°C. Unfortunately, there is significant inertia in the climate system: there is a 25 to 50 year lag between CO2 emissions and eventual warming,[25] and it is expected that 40% of the peak concentration of CO2 will remain in the atmosphere 1,000 years after the peak is reached.[26] Consequently, it is impossible to reduce temperatures quickly by reducing CO2 emissions. If the world does start to face costly warming, the international community will therefore face strong incentives to find other ways to reduce global temperatures.

The only known way to reduce global temperatures quickly and cheaply is a form of climate engineering called Solar Radiation Management (SRM), which involves cooling the Earth by reflecting sunlight back into space.[27] The most researched form of SRM involves injecting aerosols into the stratosphere.[28] Most of the evidence so far suggests that ideal SRM deployment programmes would reduce overall damages relative to an un-engineered greenhouse world.[29]

However, SRM brings its own risks. Of the currently known potential negative direct effects of SRM, only abrupt termination could plausibly bring about an existential catastrophe.[30] If a very thick stratospheric veil were deployed and SRM was suddenly terminated and not resumed within a buffer period of a few months, then there would be very rapid and damaging warming. There is some reason to think that an SRM system could be made very resilient to this *termination shock risk*, especially if the knowledge and capability for SRM deployment was widely shared, but termination shock could occur as a result of another global catastrophe, such as a global war or devastating pandemic.[31] Aside from the known risks of SRM, current climate models are imperfect, and SRM could have currently unforeseen catastrophic effects.

SRM also creates some indirect risks. Firstly, SRM could be unilaterally deployed[32] and it will be difficult to agree on a level of SRM acceptable to all regions.[33] Some nations could counter-geoengineer to reverse the effects of SRM, creating tensions and accident risk. Natural regional weather events could be incorrectly attributed to SRM, potentially leading to opposition from some parts of the world. SRM therefore poses serious international governance problems and, if mishandled, could increase the risk of political conflict.[34]

Secondly, there is a concern that SRM research or deployment could be a moral hazard by causing reduced interest in greenhouse gas mitigation.[35] If SRM does cause greenhouse gases to increase much more quickly than they would otherwise have done, this would be costly. It would be harder to keep temperature and precipitation within safe bounds, and an

increasingly thick veil would have to be deployed,[36] thereby increasing the risk of termination shock.

In this area, as for many others, the attention of researchers is of critical importance. The risks from climate change and engineering are novel phenomena and our understanding of the risks and countermeasures remains inadequate. Climate engineering, in particular, receives limited research attention.

1.1.3 Engineered pandemics

For most of human history, natural pandemics have posed the greatest risk of mass global fatalities.[37] However, there are some reasons to believe that natural pandemics are very unlikely to cause human extinction. Analysis of the International Union for Conservation of Nature (IUCN) red list database has shown that of the 833 recorded plant and animal species extinctions known to have occurred since 1500, less than 4% (31 species) were ascribed to infectious disease.[38] None of the mammals and amphibians on this list were globally dispersed, and other factors aside from infectious disease also contributed to their extinction. It therefore seems that our own species, which is very numerous, globally dispersed, and capable of a rational response to problems, is very unlikely to be killed off by a natural pandemic.

One underlying explanation for this is that highly lethal pathogens can kill their hosts before they have a chance to spread, so there is a selective pressure for pathogens not to be highly lethal. Therefore, pathogens are likely to co-evolve with their hosts rather than kill all possible hosts.[39]

Recent developments in biotechnology may, however, give people the capability to design pathogens which overcome this trade-off. Some gain-of-function research has demonstrated the feasibility of altering pathogens to create strains with dangerous new features, such as vaccine-resistant smallpox[40] and human-transmissible avian flu,[41] with the potential to kill millions or even billions of people. For an engineered pathogen to derail humanity's long-term future, it would probably have to have extremely high fatality rates or destroy reproductive capability (so that it killed or prevented reproduction by all or nearly all of its victims), be extremely infectious (so that it had global reach), and have delayed onset of symptoms (so that we would fail to notice the problem and mount a response in time).[42] Making such a pathogen would be close to impossible at present. However, the cost of the technology is falling rapidly,[43] and adequate expertise and modern laboratories are becoming more available. Consequently, states and perhaps even terrorist groups could eventually gain the capacity to create pathogens which could deliberately or accidentally cause an existential catastrophe.

1.1.4 Artificial intelligence

Currently, artificial intelligence can outperform humans in a number of narrow domains, such as playing chess and searching data. As artificial intelligence researchers continue to make progress, though, these domains are highly likely to grow in number and breadth over time.

Many experts now believe there is a significant chance that a machine superintelligence – a system that can outperform humans at all relevant intelligence tasks – will be developed within the next century. In a 2014 survey of artificial intelligence experts, the median expert estimated that there is a 50% chance of human-level artificial intelligence by 2040, and that once human-level artificial intelligence is achieved, there is a 75% chance of superintelligence in the following 30 years.[44] Although small sample size, selection bias, and the unreliability of subjective opinions mean that these estimates warrant scepticism, they nevertheless suggest that the possibility of superintelligence ought to be taken seriously.

If a superintelligence comes to exist, it will plausibly usher in economic, social, and political changes of a magnitude significantly beyond those wrought by the Industrial Revolution. While it could certainly offer many benefits, such as increased economic productivity and solutions to various technical problems, superintelligence could also be a factor in increasing existential risk.

Firstly, it could exacerbate other existential risks by destabilising political equilibria or by enabling the creation and deployment of other dangerous technologies. Secondly, it could cause grave harm through unintended consequences: the technology could be so opaque and powerful as to make it hard to ensure that it behaves in a way conducive to human good. There are a number of difficult technical problems related to the design of accident-free artificial-intelligence systems that have only recently been recognised.[45] If superintelligence comes to exist before these problems are solved then it could itself constitute an existential risk.[46]

1.1.5 Global totalitarianism

During the twentieth century, citizens of several nations lived for a time under extremely brutal and oppressive regimes.[47] Between them, these states killed more than one hundred million people, and sought total control over their citizens. Previous totalitarian states have not been particularly durable chiefly due to the problem of ensuring orderly transition between leaders, and to external competition from other more liberal and successful states. However, there is a non-negligible chance that the world will come to be dominated by one or a handful of totalitarian states. If this were to happen, external compe-

tition would no longer threaten the durability of such states to the same extent.

Moreover, improvements in certain forms of technology may make it easier for totalitarian states to maintain control, for example by making surveillance much easier. Global totalitarianism could exacerbate other existential risks by reducing the quality of governance. In addition, a long future under a particularly brutal global totalitarian state could arguably be worse than complete extinction.

1.1.6 Natural processes

As we said at the start of this section, natural existential risks appear to be less serious than anthropogenic risks. The leading natural existential risks of which we are currently aware are Near Earth Objects (asteroids and comets), super-volcanoes, and Gamma Ray Bursts. These processes have been posited as causes of the five largest mass extinctions in history.[48]

According to the US National Academy of Sciences, as a rule of thumb, Near Earth Object (NEO) impacts with a diameter of 1.5km would likely kill 10% of the world population, and the damage ramps up to the entire population for those with a diameter of 10km.[49] Due to the success of NEO tracking efforts, we can have relatively high confidence in the probability estimates of NEO strikes.[50] On average, 5km NEOs are expected to strike once every 30 million years, and 10km NEOs once every 100 million years.[51] We have discovered around 94% of nearby asteroids with a diameter of 1km or more and NASA believes all asteroids with a diameter of 10km or more have been detected,[52] and continued detection of both asteroids and comets would give us time to prepare if a large NEO were on course to hit Earth. There is at present no known feasible way to deflect NEOs with a diameter of more than a few kilometres,[53] though we might be able to develop such technology in the future.

Although they tend to receive less attention than NEOs super-volcanoes are possibly the natural existential risk that poses the highest probability of extinction. The magnitude of an eruption is measured by the Volcanic Explosivity Index (VEI). Volcanoes which are tens to hundreds of times larger than those which caused most large eruptions attain a VEI of 8 and are labelled 'super-volcanoes'.[54] Most of the damage of a super-eruption would be through a *volcanic winter*, in which the ejection of massive amounts of sulphur dioxide and smoke into the atmosphere leads to global cooling of 3-5°C for several years. Although this could kill a significant portion of the world population, it seems unlikely that it would cause extinction.[55]

Estimates of the frequency of VEI=8 eruptions vary from 30,000 years to around 130,000 years.[56] Eruptions with VEI=9 (around ten times greater in magnitude than VEI=8 eruptions) or more might cause an existential catastrophe; some experts estimate that these occur around once every 30 million years, although there is enormous uncertainty about this estimate.[57] There is little we can do to reduce super-volcano risk other than building resilience - especially by developing foods which do not depend on sunlight - and improving predictions of eruptions.

Gamma Ray Bursts (GRBs) are narrow beams of very energetic radiation probably produced by supernova explosions or mergers between compact objects such as neutron stars and black holes.[58] A sufficiently close, long and powerful GRB pointed at the Earth would chiefly do damage through massive ozone depletion leading to increased UVB radiation. In addition, large amounts of NO2 would be released into the atmosphere leading to reduced sunlight and global cooling.[59] Fortunately, potentially extinction-level GRBs are extremely rare: the mean rate is estimated to be one every 200 million years, although there is great uncertainty about this.[60] Given their frequency, they might have been responsible for previous mass extinctions.[61] In principle, we may be able to predict long GRBs,[62] but there is little we can do to prepare for them.

1.1.7 Unknown unknowns

Many of the risks discussed above were unforeseeable a few decades before they started to pose a threat. At the beginning of the 20th century, few could have anticipated that nuclear weapons, climate change, engineered pandemics, and artificial intelligence would come to be among our most severe existential risks. These risks were chiefly the products of technological and economic progress and it is inherently difficult to predict how such processes will play out. It therefore seems likely that some future existential risks, driven by the same mechanisms, are currently unknown. For example, there may be an as yet undeveloped technology which will have huge destructive power, or some way of interacting with the environment which will threaten complete ecosystem collapse.

It is of course impossible to comprehensively plan for such risks, but there are nevertheless steps we can take to reduce our vulnerability to them. Bodies tasked with horizon scanning and especially the monitoring of emerging technologies could help us to identify risks quickly as they develop. There may be generic forms of resiliency that protect against threats whose exact features we do not know, but about which we have heuristic information[63], for example by using redundant systems. Furthermore, since all existential risks, known and unknown, present a fundamentally global political challenge, greater international cooperation will reduce the threat

they pose.[64] Lastly, the significance of unknown unknowns makes it extremely important to involve research communities in efforts to address existential risk. We remain uncertain about the sources of risk and the best responses to them, and the novelty of many risks means that research work to help overcome them is a high priority.

1.2. THE ETHICS OF EXISTENTIAL RISK

In his book *Reasons and Persons*, Oxford philosopher Derek Parfit advanced an influential argument about the importance of avoiding extinction:

> *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. ... The Earth will remain habitable for at least another billion years. Civilization began only a few thousand years ago. If we do not destroy mankind, these few thousand years may be only a tiny fraction of the whole of civilized human history. The difference between (2) and (3) may thus be the difference between this tiny fraction and all of the rest of this history. If we compare this possible history to a day, what has occurred so far is only a fraction of a second.*[65]

In this argument, it seems that Parfit is assuming that the survivors of a nuclear war that kills 99% of the population would eventually be able to recover civilisation without long-term effect. As we have seen, this may not be a safe assumption – but for the purposes of this thought experiment, the point stands. What makes existential catastrophes especially bad is that they would "destroy the future," as another Oxford philosopher, Nick Bostrom, puts it.[66] This future could potentially be extremely long and full of flourishing, and would therefore have extremely large value. In standard risk analysis, when working out how to respond to risk, we work out the *expected value* of risk reduction, by weighing the probability that an action will prevent an adverse event against the severity of the event. Because the value of preventing existential catastrophe is so vast, even a tiny probability of prevention has huge expected value.[67]

Of course, there is persisting reasonable disagreement about ethics and there are a number of ways one might resist this conclusion.[68] Therefore, it would be unjustified to be overconfident in Parfit and Bostrom's argument.

In some areas, government policy does give significant weight to future generations. For example, in assessing the risks of nuclear waste storage, governments have considered timeframes of thousands, hundreds of thousands, and even a million years.[69] Justifications for this policy usually appeal to principles of *intergenerational equity* according to which future generations ought to get as much protection as current generations.[70] Similarly, widely accepted norms of sustainable development require development that meets the needs of the current generation without compromising the ability of future generations to meet their own needs.[71]

However, when it comes to existential risk, it would seem that we fail to live up to principles of intergenerational equity. Existential catastrophe would not only give future generations less than the current generations; it would give them *nothing*. Indeed, reducing existential risk plausibly has a quite low cost for us in comparison with the huge expected value it has for future generations. In spite of this, relatively little is done to reduce existential risk. Unless we give up on norms of intergenerational equity, they give us a strong case for significantly increasing our efforts to reduce existential risks.

1.3. WHY EXISTENTIAL RISKS MAY BE SYSTEMATICALLY UNDERINVESTED IN, AND THE ROLE OF THE INTERNATIONAL COMMUNITY

In spite of the importance of existential risk reduction, it probably receives less attention than is warranted. As a result, concerted international cooperation is required if we are to receive adequate protection from existential risks.

1.3.1. Why existential risks are likely to be underinvested in

There are several reasons why existential risk reduction is likely to be underinvested in. Firstly, it is a *global public good*. Economic theory predicts that such goods tend to be underprovided. The benefits of existential risk reduction are widely and indivisibly dispersed around the globe from the countries responsible for taking action. Consequently, a country which reduces existential risk gains only a small portion of the benefits but bears the full brunt of the costs. Countries thus have strong incentives to free ride, receiving the benefits of risk reduction without contributing. As a result, too few do what is in the common interest.

Secondly, as already suggested above, existential

risk reduction is an *intergenerational* public good: most of the benefits are enjoyed by future generations who have no say in the political process. For these goods, the problem is *temporal* free riding: the current generation enjoys the benefits of inaction while future generations bear the costs.

Thirdly, many existential risks, such as machine superintelligence, engineered pandemics, and solar geoengineering, pose an unprecedented and uncertain future threat. Consequently, it is hard to develop a satisfactory governance regime for them: there are few existing governance instruments which can be applied to these risks, and it is unclear what shape new instruments should take. In this way, our position with regard to these emerging risks is comparable to the one we faced when nuclear weapons first became available.

Cognitive biases also lead people to underestimate existential risks. Since there have not been any catastrophes of this magnitude, these risks are not salient to politicians and the public.[72] This is an example of the misapplication of the *availability heuristic*, a mental shortcut which assumes that something is important only if it can be readily recalled.

Another cognitive bias affecting perceptions of existential risk is scope neglect. In a seminal 1992 study, three groups were asked how much they would be willing to pay to save 2,000, 20,000 or 200,000 birds from drowning in uncovered oil ponds. The groups answered $80, $78, and $88, respectively.[73] In this case, the size of the benefits had little effect on the scale of the preferred response. People become numbed to the effect of saving lives when the numbers get too large. [74] Scope neglect is a particularly acute problem for existential risk because the numbers at stake are so large. Due to scope neglect, decision-makers are prone to treat existential risks in a similar way to problems which are less severe by many orders of magnitude. A wide range of other cognitive biases are likely to affect the evaluation of existential risks.[75]

1.3.2. The role of the international community

The international community has a crucial role to play in solving the above problems and effectively reducing existential risk. Free riding is a pervasive phenomenon, but it is especially difficult to overcome at the global level. National public goods, such as clean air, defence, and education, are provided by well-established market or centralised governmental mechanisms. At the global level, no comparable mechanisms exist. The principles of the system of international law give nation-states the right to consent to join international agreements, and all agreements are therefore voluntary.[76] Under these conditions, the incentives to free ride are powerful, and effective action depends on significant negotiation and trust. Moreover, even if international cooperation is successful, the intergenerational externalities of existential risks pose a further challenge for governance.

In spite of these barriers to political action, we know from past experience that effective provision of global and intergenerational public goods is possible. In 1989, only a few years after the discovery of significant anthropogenic ozone depletion, the Montreal Protocol came into force.[77] The treaty regulates the production of ozone depleting substances and has been ratified by nearly all states. Although there was no centralised enforcement body, and although the costs were borne by the current generation and many of the benefits will be enjoyed by future generations, nation-states have resisted the temptation to free ride. As a result, the ozone layer is predicted to fully regenerate by the middle of the century.[78] Of course, many existential risks may be much harder to resolve, and could therefore require unprecedented global cooperation.

In addition to international treaties, established supra-national organisations help to solve global collective action problems. For example, the World Health Organisation is tasked with coordinating international regulation of pandemic diseases.[79]

Anthropogenic existential risks present a novel political challenge for the international community. Technological development is making it increasingly easy for agents to bring about damaging effects with global scope. In addition, we have no track record of dealing with these technological risks, some of which could emerge very quickly and be even harder to control than nuclear weapons. In this context, international cooperation is even more essential. In the next section, we outline opportunities for the international community to address existential risk.

Endnotes – 1. An introduction to existential risks

1. This chart is adapted from Nick Bostrom, "Existential Risk Prevention as Global Priority," Global Policy 4, no. 1 (February 1, 2013): 17, doi:10.1111/1758-5899.12002.

2. Bostrom, "Existential Risk Prevention as Global Priority," 15.

3. See the discussion of differential technological development in Nick Bostrom, "Existential Risks - Analyzing Human Extinction Scenarios and Related Hazards," Journal of Evolution and Technology 9 (2002): 31–33.

4. See Toby Ord, "Will We Cause Our Own Extinction? Natural versus Anthropogenic Extinction Risks" 2014.

5. This list does not include several existential risks which have been raised but whose details remain more speculative. The ordering of the risks below does not imply a judgement about the seriousness of the risks. For a more detailed examination of these risks see Global Priorities Project and Global Challenges Foundation, "Global Catastrophic Risks 2016," 2016.

6. See Global Priorities Project and Global Challenges Foundation, "Global Catastrophic Risks 2016," chap. 2.

7. For discussion of this, see Nick Beckstead, "The Long-Term Significance of Reducing Global Catastrophic Risks," The GiveWell Blog, August 13, 2015, http://blog.givewell.org/2015/08/13/the-long-term-significance-of-reducing-global-catastrophic-risks/.

8. See Joseph Cirincione, "The Continuing Threat of Nuclear War," in Global Catastrophic Risks, ed. Nick Bostrom and Milan M. Ćirković (Oxford University Press, 2008).

9. Robock, "Nuclear Winter," 424.

10. Under New START this will be in the vicinity of the number of deployed strategic nuclear warheads owned by Russia and the USA combined.

11. See http://thebulletin.org/nuclear-notebook-multimedia.

12. For an overview of the literature see Baum, "Winter-Safe Deterrence."

13. Givewell, "Nuclear Weapons Policy," September 2015, http://www.givewell.org/labs/causes/nuclear-weapons-policy.

14. Ibid.

15. The upper limit of the likely range of warming on the high emissions pathway is 4.8°C. The IPCC does not mention the possibility that warming of around 4°C could cause human extinction. See IPCC, Climate Change 2014: Impacts, Adaptation, and Vulnerability: Summary for Policymakers (Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 2014).

16. Steven C. Sherwood and Matthew Huber, "An Adaptability Limit to Climate Change due to Heat Stress," Proceedings of the National Academy of Sciences 107, no. 21 (May 25, 2010): 9552–55, doi:10.1073/pnas.0913352107.

17. David S. Battisti and Rosamond L. Naylor, "Historical Warnings of Future Food Insecurity with Unprecedented Seasonal Heat," Science 323, no. 5911 (January 9, 2009): 240–44, doi:10.1126/science.1164363.

18. Martin L. Weitzman, "Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change," Review of Environmental Economics and Policy 5, no. 2 (July 1, 2011): 282, doi:10.1093/reep/rer006.

19. This level of warming would render most of the tropics uninhabitable and would lead to massive droughts and floods, causing unprecedented migration and probably increased conflict. See David King et al., "Climate Change–a Risk Assessment" (Centre for Science Policy, University of Cambridge, 2015), pt. 2, www.csap.cam.ac.uk/projects/climate-change-risk-assessment/.

20. Ibid., 42; Gernot Wagner and Martin L. Weitzman, Climate Shock : The Economic Consequences of a Hotter Planet (Princeton: Princeton University Press, 2015), 50.

21. Weitzman, "Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change."

22. Ibid.

23. Wagner and Weitzman, Climate Shock, 2015, chap. 3.

24. Weitzman, "Fat-Tailed Uncertainty in the Economics of Catastrophic Climate Change," 279.

25. James Hansen et al., "Earth's Energy Imbalance: Confirmation and Implications," Science 308, no. 5727 (June 3, 2005): 1431–35, doi:10.1126/science.1110252.

26. Susan Solomon et al., "Irreversible Climate Change due to Carbon Dioxide Emissions," Proceedings of the National Academy of Sciences 106, no. 6 (February 10, 2009): 1704–9, doi:10.1073/pnas.0812721106.

27. One possible option would be to reduce methane emissions, which accounts for half the total warming of CO2, and which is removed from the atmosphere after around 12 years. See Climate Change Division US EPA, "Methane Emissions," Overviews & Factsheets, https://www3.epa.gov/climatechange/ghgemissions/gases/ch4.html.

28. Shepherd, Geoengineering the Climate, chap. 3.

29. For a good overview see Oliver Morton, The Planet Remade: How Geoengineering Could Change the World (London: Granta, 2015), chap. 4.

30. Seth D. Baum, Timothy M. Maher, and Jacob Haqq-Misra, "Double Catastrophe: Intermittent Stratospheric Geoengineering Induced by Societal Collapse," Environment Systems & Decisions 33, no. 1 (January 8, 2013): 168–80, doi:10.1007/s10669-012-9429-y.

31. We are very grateful to Andy Parker for helpful discussion of this point.

32. Edward A. Parson, "Climate Engineering in Global Climate Governance: Implications for Participation and Linkage," Transnational Environmental Law 3, no. 1 (April 2014): 89–110, doi:10.1017/S2047102513000496.

33. Morton, The Planet Remade, 2015, chap. 4.

34. For an overview see Paul Nightingale and Rose Cairns, "The Security Implications of Geoengineering: Blame, Imposed Agreement and the Security of Critical Infrastructure," Climate Geoengineering Governance Working Paper Series (Tech. Rep., School of Business, Management and Economics, Univ. of Sussex, 2015).

35. For an excellent discussion of the moral hazard worry see David R. Morrow, "Ethical Aspects of the Mitigation Obstruction Argument against Climate Engineering Research," Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 372, no. 2031 (December 28, 2014): 20140062, doi:10.1098/rsta.2014.0062.

36. Morton, The Planet Remade, 2015, chap. 4.

37. See the discussion in Global Priorities Project and Global Challenges Foundation, "Global Catastrophic Risks 2016," chap. 2.

38. Katherine F. Smith, Dov F. Sax, and Kevin D. Lafferty, "Evidence for the Role of Infectious Disease in Species Extinction and Endangerment," Conservation Biology 20, no. 5 (October 1, 2006): 1349–57, doi:10.1111/j.1523-1739.2006.00524.x.

39. Matthew C. Fisher et al., "Emerging Fungal Threats to Animal, Plant and Ecosystem Health," Nature 484, no. 7393 (April 12, 2012): 186–94, doi:10.1038/nature10947.

40. Christopher F. Chyba and Alex L. Greninger, "Biotechnology and Bioterrorism: An Unprecedented World," Survival 46, no. 2 (2004): 143–162.

41. Masaki Imai et al., "Experimental Adaptation of an Influenza H5 HA Confers Respiratory Droplet Transmission to a Reassortant H5 HA/H1N1 Virus in Ferrets," Nature 486, no. 7403 (June 21, 2012): 420–28, doi:10.1038/nature10831.

42. Martin J. Rees, Our Final Century: Will Civilisation Survive the Twen-

ty-First Century? (London: Arrow, 2004), 45–47.

43. Robert Carlson, "The Changing Economics of DNA Synthesis," Nature Biotechnology 27, no. 12 (December 2009): 1091–94, doi:10.1038/nbt1209-1091.

44. Vincent C. Müller and Nick Bostrom, "Future Progress in Artificial Intelligence: A Survey of Expert Opinion," Fundamental Issues of Artificial Intelligence (2016): 553-70.

45. Dario Amadei et al., "Concrete Problems in AI Safety" (2016). https://arxiv.org/pdf/1606.06565v2.pdf.

46. For an extended discussion of how to control an MASI see Nick Bostrom, Superintelligence: Paths, Dangers, Strategies (Oxford: Oxford University Press, 2014).

47. For an overview see Bryan Caplan, "The Totalitarian Threat," in Global Catastrophic Risks, ed. Nick Bostrom and Milan M. Ćirković (Oxford: Oxford University Press, 2008).

48. Arnon Dar, "Influence of Supernovae, Gamma-Ray Bursts, Solar Flares, and Cosmic Rays on the Terrestrial Environment," in Global Catastrophic Risks, ed. Nick Bostrom and Milan M. Ćirković (Oxford: Oxford University Press, 2008).

49. National Research Council (U. S.). Committee to Review Near-Earth-Object Surveys and Hazard Mitigation Strategies, Defending Planet Earth: Near-Earth Object Surveys and Hazard Mitigation Strategies (Washington, DC: National Academies Press, 2010), 23.

50. Alan Harris, "What Spaceguard Did," Nature 453, no. 7199 (June 26, 2008): 1178–79, doi:10.1038/4531178a.

51. National Research Council (U. S.), Defending Planet Earth, 19.

52. Dr Alan Harris, personal email correspondence, 11th July 2016.

53. Ibid., chap. 5.

54. Michael Rampino, "Super-Volcanism and Other Geophysical Processes of Catastrophic Import," in Global Catastrophic Risks, ed. Nick Bostrom and Milan M. Ćirković (Oxford: Oxford University Press, 2008).

55. Rampino argues that a VEI = 8 event would produce the same damage as a 1.5km asteroid, which is unlikely to produce extinction. Ibid., 215. For some discussion of the uncertainty in the literature see Open Philanthropy Project, "Large Volcanic Eruptions," June 2013, http://www.openphilanthropy.org/research/cause-reports/volcanoes.

56. W. Aspinall et al., "Volcano Hazard and Exposure in GFDRR Priority Countries and Risk Mitigation Measures," Volcano Risk Study 0100806-00-1-R, 2011, 15; Susan Loughlin et al., Global Volcanic Hazards and Risk (Cambridge University Press, 2015), 97.

57. Aspinall et al., "Volcano Hazard and Exposure in GFDRR Priority Countries and Risk Mitigation Measures," 15.

58. Brian C. Thomas, "Gamma-Ray Bursts as a Threat to Life on Earth," International Journal of Astrobiology 8, no. 3 (2009): 183.

59. Ibid.

60. Tsvi Piran and Raul Jimenez, "Possible Role of Gamma Ray Bursts on Life Extinction in the Universe," Physical Review Letters 113, no. 23 (December 5, 2014): 231102, doi:10.1103/PhysRevLett.113.231102.

61. A.l. Melott et al., "Did a Gamma-Ray Burst Initiate the Late Ordovician Mass Extinction?," International Journal of Astrobiology 3, no. 1 (January 2004): 55–61, doi:10.1017/S1473550404001910.

62. Brian Thomas, personal correspondence, 8th July 2016.

63. Seth D. Baum, 2015. Risk and resilience for unknown, unquantifiable, systemic, and unlikely/catastrophic threats. Environment, Systems, and Decisions, vol. 35, no. 2 (June), pages 229-236.

64. We return to this in section 2.3.

65. Derek Parfit, Reasons and Persons (Oxford: Oxford University Press, 1984), pp. 453-454.

66. Bostrom, "Existential Risk Prevention as Global Priority", 15. Nick Bostrom, "Astronomical Waste: The Opportunity Cost of Delayed Technological Development," Utilitas 15, no. 3 (November 2003): 308–314, doi:10.1017/S0953820800004076; Bostrom, "Existential Risk Prevention as Global Priority"; Nicholas Beckstead, "On the Overwhelming Importance of Shaping the Far Future" (Rutgers University-Graduate School-New Brunswick, 2013), https://rucore.libraries.rutgers.edu/rutgers-lib/40469/.

67. Bostrom, "Existential Risk Prevention as Global Priority," 18–19.

68. For in-depth discussion of some possible counter-arguments see Beckstead, "On the Overwhelming Importance of Shaping the Far Future," chaps. 3–8.

69. L. H. Hamilton et al., "Blue Ribbon Commission on America's Nuclear Future: Report to the Secretary of Energy" (Blue Ribbon Commission, 2012), 90.

70. This has been agreed to by signatories to the International Atomic Energy Agency's Joint Convention on nuclear waste management. See International Atomic Energy Agency, "Joint Convention on the Safety of Spent Fuel Management and on the Safety of Radioactive Waste Management," December 24, 1997, chap. 3, https://inis.iaea.org/search/search.aspx?orig_q=RN:36030798.

71. The most widely accepted definition of 'sustainable development' originated with the Brundtland Report. See Gru Brundtland et al., Report of the World Commission on Environment and Development: Our Common Future (Oxford University Press, 1987).

72. For a discussion of the cognitive biases affecting judgements of global risks see Eliezer Yudkowsky, "Cognitive Biases Potentially Affecting Judgment of Global Risks," in Global Catastrophic Risks, ed. Nick Bostrom and Milan M. Ćirković (Oxford: Oxford University Press, 2008).

73. William H. Desvousges et al., "Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy," Research Triangle Institute Monograph, 1992. For other examples of scope neglect see Daniel Kahneman et al., "Economic Preferences or Attitude Expressions?: An Analysis of Dollar Responses to Public Issues," Journal of Risk and Uncertainty 19, no. 1–3 (1999): 203–35.

74. Fetherstonhaugh D, Slovic P, Johnson SM, Friedrich J. Insensitivity to the Value of Human Life: A Study of Psychophysical Numbing. J Risk Uncertain. 1997;14:283–300.

75. Yudkowsky E. Cognitive Biases Potentially Affecting Judgment of Global Risks. In: Bostrom N, Cirkovic M, editors. Global Catastrophic Risks. Oxford University Press; 2008. p. 91–119.

76. For a discussion of free riding in climate treaty negotiations see William Nordhaus, "Climate Clubs: Overcoming Free-Riding in International Climate Policy," The American Economic Review 105, no. 4 (2015): 1339–1370.

77. Scott Barrett, Why Cooperate? : The Incentive to Supply Global Public Goods (Oxford: Oxford University Press, 2007), chap. 3.

78. Ibid., 83.

79. See WHO, "International Health Regulations: Support to Global Outbreak Alert and Response, and Building and Maintaining National Capacities," 2015, http://apps.who.int/iris/bitstream/10665/199747/1/WHO_HSE_GCR_2015.7_eng.pdf.

2. Recommendations

Having outlined the major existential risks, we will now outline the interventions that we regard as the most promising ways to reduce existential risk. This is based on an initial overview survey, the details of which can be found in the appendix.

These are early days for work to reduce existential risk. In the course of this survey we also identified other promising interventions which did not make it to this final list but have significant potential. We outline those proposals in the appendix and outline the methodology by which these proposals were selected. We also remain confident that there are very valuable ideas out there which have not yet been considered.

Our current recommendations are to:

1. Develop governance of Solar Radiation Management research
2. Establish scenario plans and exercises for severe engineered pandemics at the international level
3. Build international attention to and support for existential risk reduction

2.1. DEVELOP GOVERNANCE OF SOLAR RADIATION MANAGEMENT RESEARCH

The international community should increase diplomatic awareness of the issues raised by Solar Radiation Management, with a view to developing appropriate codes of conduct, governance and/or regulation.

2.1.1 Current context

Solar Radiation Management (SRM) is the application of techniques to reflect sunlight and reduce global temperatures. Currently, the leading proposed approach is the use of stratospheric sulphate aerosols. Improved SRM research and governance of this research would help to reduce the risks both from climate change and from SRM itself.

Although it may seem remote now, there are likely to be strong incentives to use SRM in the future, even if research into it has not progressed much further. As we argued in Section 1, as the future costs of climate change begin to be felt, some or all of the international community will face strong pressures to use SRM because it is the only known way to reduce global temperatures quickly and cheaply. This suggests that further SRM research would be valuable, provided steps are taken to reduce potential moral hazard effects on mitigation.[1] All of the major reports on SRM call for more research and for governance of this research where appropriate.[2] Further research is particularly important, because the premature rejection of climate engineering could be as risky for our climate as premature use in the context of ongoing climate change.

SRM research is currently in its infancy, currently conducted mostly through computer modelling, but there are increasing calls within the research community for outdoor field testing.[3] If conducted at large scale, these tests could create risks of transboundary harm.[4] This means that formalised governance arrangements may be necessary, whether through voluntary scientific codes of conduct or through international treaties.

Geoengineering governance is likely to fall under the purview of the United Nations Environment Programme. However, it is not yet clear how present international law will be applied to SRM in practice: There is no present international treaty explicitly designed to govern SRM research or deployment, and existing instruments may not be well-suited to the novel challenges posed by SRM.[5]

Some NGOs have argued that the Convention on Biological Diversity (CBD) should take an overarching role in geoengineering deployment governance,[6] but this approach may not work for a number of reasons. Firstly, the mandate of the CBD is limited to addressing threats to biodiversity, which is only one of many considerations regarding the benefits and risks of SRM. Secondly, the US is not a signatory to the CBD.

In addition, this kind of early approach may fall foul of the technology control dilemma: It is impossible to know in advance how SRM technology will turn out. If SRM deployment governance is developed too early, it is likely to be ill-suited to the technologies it is supposed to control.[7] However, as discussed below, there are also reasons in favour of developing governance early. It may be that informal but clear and explicit agreements between researchers would be more valuable and adaptable than formal instruments.

2.1.2 Proposed intervention

We propose that members of the international governance community should increase efforts to build awareness of the policy issues raised by SRM, with a view to eventually developing instruments for the appropriate governance and regulation of SRM research.

This governance must be appropriate for the research activity targeted. A legal regime for computer modelling would be regulatory overkill,[8] but international governance could be required for some field

TYPES OF INTERVENTIONS TO REDUCE EXISTENTIAL RISK

Throughout the process, we employed multiple classification systems to help us cover a wider range of possibilities. The three most useful classifications were:

Prevention vs. response vs. endurance vs. recovery

Interventions can focus on different time points in the development of a risk:

- Preventative interventions reduce the likelihood of the event occurring, or reduce the likelihood that the risk becomes existential (for non-existential risks, this second aspect is often called mitigation).
- Interventions that improve response capacity help people to manage the immediate impact of an event.
- Interventions that improve endurance make it easier for people to survive the aftermath of an event while the environment becomes less adverse.
- Interventions that improve recovery make it easier for survivors to rebuild a flourishing civilisation.

Strategies that focus on prevention are often appealing because usually they mean significantly less total harm (depending on the cost of the preventative strategy). Moreover, because many of the causal mechanisms underlying existential risks are in their early stages related to more frequent risks, some prevention approaches can be easily integrated with work that addresses disaster risks with more established support.

Response, endurance, and recovery are often thought of as aspects of resilience.

Cross-cutting vs. risk-specific

Some interventions address specific risks (e.g., geoengineering governance) while others help with multiple risks (e.g., food and medical stockpiles). Both types are important and the ideal global strategy probably employs a mixture of the two. However, cross-cutting work is likely to end up more neglected, because it can require coordination between multiple areas of governance which are typically independent.

In general, prevention work tends to be more risk-specific because it engages with a specific causal process that creates a risk. In contrast, resilience strategies tend to be more cross-cutting because the consequences of many risks are similar.

Direct vs. capacity-building

Some interventions will directly reduce an existential risk, while others increase the global community's capacity to reduce them in the future. This distinction runs along a spectrum. Towards the direct end are activities like directly dismantling nuclear stockpiles, and at the other end are activities like advocating for increased research into existential risk reduction. In between are a range of strategies with varying levels of directness, such as advocating for nuclear stockpile reduction or offering training in safe nuclear disarmament.

Capacity building strategies often include research into risk reduction techniques and policies, raising awareness of and drawing attention to existential risk, and building organisations and institutions which will work to reduce existential risk, directly or indirectly. These strategies are often more appropriate for emerging or poorly-understood risks, where the specific steps required to address them are not yet known. They also offer the possibility for small group to improve the efficiency of much larger pools of resources, and they can therefore potentially be more cost-effective.

In the long run, of course, it is direct work which will make the difference, and capacity-building is valuable only because it enables future direct work. Right now, direct work is especially feasible for the better understood risks such as climate change and nuclear warfare, as well as to demonstrate the plausibility (or proof of concept) of emerging risk reduction strategies.

tests, such as those that pose a risk of transboundary harm (though it is worth noting that some outdoor field testing would not pose the risk of such harm).[9]

As geoengineering researchers Professors Ted Parson and David Keith argue, *"progress on research governance is needed that advances four aims: (i) letting low-risk scientifically valuable research proceed; (ii) giving scientists guidance on the design of socially acceptable research; (iii) addressing legitimate public concern about reckless interventions or a thoughtless slide from small research to planetary manipulation; and (iv) ending the current legal void that facilitates rogue projects".*[10]

Geoengineering governance researchers have proposed principles which could help guide policymakers in building SRM governance.

(a) The Oxford Principles

Following the publication of the Royal Society report on geoengineering, Rayner et al developed a list of five high-level principles to guide geoengineering governance.[11]

Principle 1: Geoengineering to be regulated as a public good.

Principle 2: Public participation in geoengineering decision-making.

Principle 3: Disclosure of geoengineering research and open publication of results.

Principle 4: Independent assessment of impacts.

Principle 5: Governance before deployment.

The key to implementation of the Oxford Principles is the creation of research protocols for each stage of the development of geoengineering technology. The Oxford Principles are neutral on whether governance should be imposed at the nation-state level or whether a voluntary code of conduct among scientific bodies would be sufficient.

(b) The Geoengineering Governance Research Project draft Code of Conduct

The Oxford Principles are quite high level and leave a lot of room for interpretation. The draft Code of Conduct developed by the Geoengineering Governance Research Project offers more concrete guidance.[12] Draft Article 6, which recommends the development of a specific international liability and compensation scheme for those who may suffer adverse consequences from geoengineering, may be particularly relevant for the diplomatic community. However, the Code of Conduct is still in its draft stage, and so would benefit from discussion by key stakeholders, including scientists, diplomats, and policymakers.

We remain neutral on whether it is yet appropriate to develop governance of SRM deployment, as opposed to research, as there is expert disagreement on this issue. On the one hand, some argue that it would be premature to develop SRM governance due to the technology control dilemma,[13] and due to the fact that developing this governance could be a very lengthy process and could unreasonably delay valuable SRM research.[14] Others argue that large-scale field testing should not be done until there is at least some governance in place to cover the later deployment phase.[15] The reason for this is that if we focus only on research governance and leave implementation governance for later, we risk arriving at the stage where implementation becomes a real concern before we have appropriate governance in place.

It may be that the priority for governance should be support for openness and international collaboration, for example in the form of joint international projects. Since much of the risk from stratospheric sulphates comes from the possibility of termination shock, this risk is reduced if a broad range of countries possess the capability to maintain climate engineering programmes. However, the coordination issues surrounding governance are likely to become more complicated as the number of parties able to engineer the climate grows.

2.1.3 Impact of the intervention

Our proposal only pertains to SRM research, rather than SRM deployment. Outdoor SRM research is unlikely to pose an existential risk. The value of SRM research governance stems firstly from allowing us to learn more about a technology which could help reduce climate change risk in the future, and secondly from starting a process which would make adequate governance of SRM deployment more likely in the future. Overall, if done appropriately, the intervention would help to reduce the existential risks from climate change and SRM, which plausibly comprises a non-negligible fraction of overall existential risk.

That said, there are several ways in which the intervention could have zero or negative impact. Firstly, there is a chance that research governance and even the mere discussion of it could create moral hazard by reducing our willingness to mitigate greenhouse gases. However, it is unclear whether these moral hazard effects would occur, and steps can and should be taken to reduce this risk, for example through a publicised agreement that SRM does not justify the failure to mitigate. Moreover, if there were safe and effective SRM techniques it might be rational to plan to make use of them in some circumstances.

Secondly, there is a risk that regulation will inappropriately constrain valuable research, or will fail to prevent unnecessarily risky projects. If governance

follows the aforementioned principles, these possibilities should be avoided.

2.1.4 Ease of making progress

As we have said above, present international law places no control on SRM research and deployment. There have been calls from within the research community for increased research and governance of research, but the community is quite small at the moment – according to David Keith and Andy Parker, as of 2013, it received only around $11m in research funding per year.[16] This suggests that the area is not overcrowded and that additional political work has a relatively high chance of impact.

If the intervention is to gain sufficient political momentum, it will require support from key stakeholders in the scientific community and environmentalist NGOs. It is worth noting that many such people are opposed to SRM. As a result, they may be supportive of research governance in principle, as it would help to prevent overly risky SRM research. However, others may be particularly opposed to it on the basis of moral hazard concerns.

2.2. ESTABLISH SCENARIO PLANS AND EXERCISES FOR SEVERE ENGINEERED PANDEMICS AT THE INTERNATIONAL LEVEL

The international community should increase the level of scenario planning for pandemic preparedness, focused on non-naturally occurring pandemics including those engineered to have characteristics that make them difficult to manage and contain.

2.2.1 Current context

Pandemic management is challenging. An outbreak of Ebola in Western Africa in 2015 highlighted the need for significant improvements in outbreak response, especially where international coordination is needed to address expanding outbreaks in resource-poor health systems, and the international health community is already reacting to the shortcomings exposed by the outbreak.

Despite the challenges posed by Ebola, the disease is comparatively containable since transmission requires direct contact with bodily fluids. Even natural pathogens, such as zoonotic influenza, could pose a much greater international health risk. Pathogens which are deliberately engineered to make pandemic management approaches less effective, and are deployed so as to maximise the harm caused, could pose an unprecedented challenge. At the moment, very little international work is being done to plan for such scenarios.

Since the end of the Second World War, a number of states have pursued biological weapons programmes, and various terrorist groups have tried to procure biological weapons.[17] However, although some currently-available biological weapons could have truly catastrophic consequences, none are sufficiently dangerous to pose an existential threat. But as we argued in Section 1, engineered pandemics will present an increasing existential risk over the coming decades as biotechnology improves and scientific expertise becomes more widespread. Indeed, many experts have argued that they will eventually become one of the most severe existential risks.[18]

The responsibilities of the international community

National governments currently take responsibility for major disease outbreaks, along with a range of international organisations and treaties.[19] The following international bodies and agreements have an important role to play in ensuring preparedness for and response to engineered disease outbreaks.

1. **The United Nations**
   The UN Secretary General has the right to investigate the use of chemical and biological weapons. The Ebola emergency response was the UN's first ever health mission.

2. **The Biological Weapons Convention (BWC)**
   Signed by 175 State Parties, the BWC prohibits the production and use of biological weapons. The BWC also places obligations on states in the event of the use of a biological weapon: "Each State Party to this Convention undertakes to provide or support assistance, in accordance with the United Nations Charter, to any Party to the Convention which so requests, if the Security Council decides that such Party has been exposed to danger as a result of violation of the Convention" (Article VII).

3. **The World Health Organisation**
   The WHO manages public health emergencies of international concern, including both natural and engineered pandemics. For the last decade, the WHO's revised International Health Regulations (IHR) have been the legally binding governing framework for global health security.[20] The IHR states the following regarding deliberately produced disease outbreaks: "If a State Party has evidence of an unexpected or unusual public health event within its territory, irrespective of origin or source, which may constitute a public health emergency of international concern, it shall provide WHO with all relevant public health information" (Article 7). The WHO's Global Outbreak Alert and Response Network (GOARN) is a technical collaboration

of existing institutions that pool resources for the rapid identification of and response to outbreaks of international importance. In the event of the intentional release of a biological agent, GOARN would be vital for effective international containment efforts.[21] The WHO can also draw on support from its collaborating centres and laboratory networks, such as the Global Influenza Surveillance and Response System. Finally, as part of its biorisk management programme, the WHO also provides guidance on laboratory biosafety and security. This guidance would be relevant for reducing the risk of both accidental and deliberate release of pathogens from laboratories.[22]

4. **The World Organization for Animal Health**
   The World Organization for Animal Health (OiE) is an international organisation promoting animal disease control. This is important as 80% of agents with bioterror potential have zoonotic (animal-based) origin. The OiE's Biothreat Reduction Strategy explicitly addresses deliberate disease outbreaks,[23] and its World Animal Health Information System helps with disease surveillance and response. Like the WHO, the OiE also provides guidance on laboratory biosafety and biosecurity.[24]

5. **The Food and Agriculture Organization**
   The Food and Agriculture Organization (FAO) is a UN agency which works to improve global food security. It provides guidance for states on preparing for pandemics that could affect the animal population.[25]

6. **Interpol**
   As the world's largest international police organisation, Interpol enables collaboration between different police forces, and could be heavily involved in the response to a bioterror attack.

Existing scenario planning efforts

Pandemic scenario plans determine how actors will respond to different pandemic scenarios. Pandemic scenario exercises – sometimes called simulation exercises – put these plans into practice and ensure that the plans are operable in a crisis. Scenario exercises encourage key actors to consider how they would respond in a range of different scenarios. These exercises typically come in three forms.[26]

1. **Table top exercises**
   The major stakeholders are assembled to talk through a scenario, describing the actions that would be taken at each point.

2. **Functional exercises**
   Participants actually complete certain actions while working through the provided scenario. This type of exercise usually focuses on the coordination of multiple functions or organisations. Functional exercises strive for realism, short of actual deployment of equipment and personnel.

3. **Full-scale exercises**
   An emergency event is simulated as closely to reality as possible. This type of exercise involves all the named responders in the plan, and requires deployment of personnel and equipment.

Tabletop exercises are the most common form of exercise, and exercise programmes should typically start with table top exercises.[27]

The WHO, the OiE, the FAO, and Interpol all carry out scenario planning and provide scenario planning resources and guidance for states and other bodies. From the point of view of reducing existential risk, current efforts have a number of important features:

1. **Plans predominantly focus on natural disease outbreaks.** Most disease outbreak scenario plans and exercises are currently aimed at natural disease outbreaks.[28]

2. **Plans that do focus on unnatural disease outbreaks tend not to focus on pathogens which have been engineered to have dangerous new properties.** The WHO, the OiE, and Interpol have plans in place for deliberate disease outbreaks, but these tend to involve known pathogens, such as plague and smallpox, which have not been engineered to have new and dangerous properties.[29]

3. **Plans are predominantly nationally or regionally focused.** Scenario plans and exercises are chiefly nationally or regionally focused, rather than internationally focused. However, particularly in the wake of growing concern about terrorism, bioterror scenario exercises have now been performed, involving a range of international actors.[30]

2.2.2 Proposed intervention

We propose that scenario plans for very severe engineered pandemics be established at the international level, and that these plans be tested in exercises involving a range of international actors. This would

enable relevant actors to identify where more expertise is needed, where prior agreements are required for swift response, and where communication channels must be strengthened.

The focus of current scenario planning on natural pandemics at present is understandable because they pose a more immediate threat. However, some attention should be shifted towards engineered disease outbreaks because even now they pose some risk and they will pose a greater threat in the future. Moreover, insofar as states and international organisations do consider unnatural disease outbreaks, they ought to consider the implications of pathogens engineered to have new and dangerous properties. Such pathogens could be deployed deliberately or as the result of an accident at a laboratory involved in biological research, and in either case they could result in extremely severe pandemics. Although these pandemics are relatively unlikely, the scale of their impact justifies taking steps to reduce the risk. It is important for scenario plans to have a truly international focus, and to include a range of relevant actors and institutions across the public health and security communities, such as the WHO, the OiE, Interpol, and States Parties to the BWC.

Many of the governance challenges posed by engineered pathogens are similar to those posed by natural pathogens.[31] For example, neither natural nor engineered pathogens respect national borders, and international cooperation is therefore essential to limit the risks they pose. However, engineered pathogens also present some unique challenges. Firstly, preparedness for and response to engineered disease outbreaks must involve significant collaboration between the public health community and the security community. Secondly, ascertaining the source of the outbreak – whether, for example, it is a deliberate bioterror act or a laboratory accident – will have a great bearing on the appropriate response. Thirdly, depending on the properties of the engineered pathogen, standard countermeasures may be ineffective. Deliberately distributed pathogens are more likely to originate in multiple locations at the same time, making it harder to keep an outbreak regional. Aware of some of these novel potential problems, some states are now trying to determine how the international response to the Ebola outbreak would have been different if the outbreak had been deliberately caused.[32]

2.2.3 Impact of the intervention

The chief impact of the proposed intervention would be to raise awareness about the level of preparedness for emerging biotechnological risks. It could help to identify areas in which major reform is needed, as well as areas in which substantial progress could be made through other routes, such as issuing guidance, establishing better communication and information exchange systems, or running particular training programmes. This could help to improve national and international planning and cooperation, and to shift resources and attention towards reducing these risks. A similar mechanism appears to have worked in other cases. For example, partly as a result of the poor preparedness highlighted by a smallpox outbreak scenario exercise, in 2002 the US government invested heavily in smallpox countermeasures and bought enough vaccine to vaccinate every US citizen.[33]

The intervention may add further support to ongoing efforts to improve pandemic preparedness by enhancing formal international organisations, institutions and treaties. For example, it may encourage the international community to further support efforts, such as the Global Health Security Agenda, to improve the national core health capacities of low and middle income countries, as required by the IHR.[34] In addition, it might increase support for governance reforms of the WHO, such as the establishment of a Centre for Emergency Preparedness and Response which integrates and strengthens all of the WHO's preparedness, response, and humanitarian activities. The establishment of this centre was recommended by all four commissions on the Ebola crisis.[35] On the security side, further support may be added to ongoing efforts to operationalise Article VII of the BWC. This topic has received significant attention in the intersessional work of the BWC, and is likely to be discussed at the BWC Eighth Review Conference at the end of the year.[36]

2.2.4 Ease of making progress

Staging an international tabletop exercise with a range of international actors would require widespread support, but it appears to be achievable. It may be easiest for such an arrangement to take place at a regional level or between countries which already coordinate extensively, such as the EU or the Five Eyes intelligence partnership.

However, there are numerous political, diplomatic, and financial barriers to major reform of the key instruments and institutions, such as the BWC, the WHO and its IHR. For example, there was political controversy about linking public health and security in the IHR.[37] In general, revising any international treaty is likely to be a lengthy process involving years of negotiation.[38] This may be exacerbated by the difficulty of sharing potentially sensitive health data, although individual level data is unlikely to be needed.

One way to make progress without altering the formal mandate of governance institutions would be to create a core group of states and interested bodies

WORLD BANK PANDEMIC EMERGENCY FINANCING FACILITY

In May 2016, the World Bank announced the creation of a new Pandemic Emergency Financing Facility (PEF), a fast-disbursing global financing mechanism designed to help low income countries respond to outbreaks of diseases with pandemic potential.[39] It was created following the response to the 2014 Ebola outbreak in West Africa, which is widely thought to have been inadequate.[40] Significant international funds were not mobilised against Ebola until October 2014, by which time the number of cases had increased tenfold from July.[41]

The PEF only covers low income countries and includes a $500m insurance component as well as a replenishable $50-100m cash component. The insurance component covers outbreaks of infectious diseases most likely to cause major epidemics, including pandemic influenza, SARS, MERS, and Ebola. Insurance premiums are funded by donor countries, and coverage is provided by resources from the reinsurance market along with the proceeds of catastrophe bonds (index-linked securities which lose their face value if a catastrophe hits – these bonds provide a way to transfer risk from reinsurance companies to investors).

To complement the insurance window, based on financial instruments like catastrophe bonds, a the facility will also set aside cash to provide more flexible funding to address a larger set of emerging pathogens which may not meet the criteria for the insurance window.

Pandemic insurance and existential risk

As we argued in Section 1, it appears to be very unlikely that natural pandemics could bring about human extinction. However, engineered pandemics are a serious existential risk, and the PEF may indirectly help to reduce this risk. Firstly, the cash component of the PEF, or the insurance component (if the parametric trigger – which determines whether the catastrophe bonds or insurance payments trigger - is defined sufficiently broadly) could be deployed to help low income countries deal with an engineered pandemic outbreak. Secondly, if appropriately designed, the PEF could incentivise improvements in health systems and pandemic response planning, which would reduce the threat from engineered pandemics. Designing insurance systems to achieve these aims is challenging, but one possible approach would be to make coverage or the cost of the premiums conditional on certain health system improvements.[42] In addition to these indirect effects, the PEF sets a precedent for further global cooperation on major global risks, whether through insurance mechanisms or otherwise, and illustrates how improving the capacity of poor countries to deal with global risks is in everyone's interest.

However, in spite of its merits, if the PEF is to have a serious effect on pandemic risk, its funding needs to be appropriate to the scale of the risk. At present, the funds committed to the PEF appear small relative to the scale of investment required. To give an idea of scale: the response and recovery to the Ebola crisis alone ended up costing $7bn,[43] and as of 2014, the costs of financing country system investments and operations were an estimated $3.4 billion every year.[44] Therefore, we recommend that the international community increase funds available for the PEF, or for rapid access cash set aside for the WHO.

Furthermore, improving the financing of response to catastrophes is just one way to reduce the risk of extremely severe engineered pandemics. Perhaps most importantly, high income countries should provide the financial and technical support required for lower income countries to improve their core health capacities in line with WHO's International Health Regulations.[45]

that could work on engineered pandemics parallel to the existing activities of these institutions, and then report back in formal meetings. This would help to bring attention to emerging biotechnology threats while avoiding some of the usual diplomatic barriers. Furthermore, as mentioned in the previous subsection, there are other ways in which progress can be made without potentially intractable major reform, such as through improved guidance.

One sensitivity which may affect the process is that joint scenario planning may require transparency about existing biosecurity capabilities. As a result, thorough scenario planning may only be possible in groups of nations which share deep strategic trust.

2.3. BUILD INTERNATIONAL ATTENTION TO AND SUPPORT FOR EXISTENTIAL RISK REDUCTION

Increase the attention given to and support for existential risk reduction by institutions and individuals within the international community.

2.3.1 Current context

Reducing existential risk is likely to require extensive international co-ordination, co-operation, and action. Some approaches can succeed at a purely national or regional level, or within particular research communities. However, because most existential risks are essentially transnational, the international community has a large role to play.

As discussed in the rest of this report, there are several actions the international community could take to reduce existential risk either immediately, such as scenario planning for pandemics, or in the near future, such as creating a geo-engineering research governance. However, similar type of international action typically requires a high level of agreement and buy-in from decision-makers in nation-states and international institutions. This level of agreement in turn relies on the level of international attention to and support for a topic.

We can see this in the history of international action to prevent climate change. Expert debate over anthropogenic climate change recognised a serious risk in the 1950s. Although individual nations and regional groups became more aware of the issue, it did not receive significant international attention for many decades: the Intergovernmental Panel on Climate Change was not established until 1988.

Some nations and regions have taken big steps to unilaterally reduce their emissions, but many options require international coordination and mutual support. It may be that 2015's Conference of Parties of the UNFCCC[46] in Paris represents such a commitment to collective action, enabling bolder steps to be taken to address climate change. This process, lengthy though it was, may present a roadmap towards the establishment of existential risk as a priority for the international community.

Unfortunately, at the moment most international decision-makers are not particularly well-aware of existential risk. Attention is paid to some individual risks which are potentially existential – for example pandemics or nuclear warfare. However, there is relatively little international attention paid to the most extreme forms of these risks, or awareness that actions to prevent one kind of risk might also help reduce other risks. As a result, a great deal of work is done in isolation, with comparatively little effort made to communicate with other organisations with similar goals.

Although attention and support has been increasing, there is still much to be done to raise the profile of existential risk. In particular, it could be extremely beneficial to focus on building connections between communities and consolidating existing knowledge and efforts. Some of this work might pay off swiftly, while other parts could lay the foundation for work in future decades – much as the Paris Agreement depended on the foundations established by campaigns begun in the previous century.

It is important to start building international attention to existential risk early on because it can take a great deal of time to convene support and initiate action on large and complex issues. Even for extremely contained issues, building attention and support can take many years. For example, the Protocol on Blinding Laser Weapons represented a comparatively modest change to humanitarian law, but came into force 12 years after it was first raised by a member state in 1986 and has still only been signed by 107 states 30 years on. We might expect that more contentious issues which have greater implications for society broadly would take significantly longer.

In a positive development, the United Nations has made significant commitments to future generations. UNESCO, for example, acknowledged in its 1997 Declaration on the Responsibilities of the Present Generations towards Future Generations that "at this point in history, the very existence of humankind and its environment are threatened" and that "present generations should strive to ensure the maintenance and perpetuation of humankind". However, there is more to be done. Although the UN's Sendai Framework for Disaster Risk Reduction was created to guard against both small-scale and large-scale risks, it does not discuss existential risks. Disaster risk management must incorporate existential risks, or else a new field of risk management will be required if those risks are to be addressed in a coherent way.

For some of the risks we consider here, awareness is limited. Even discussions about climate change, for example, rarely acknowledge the risks from catastrophic forms of climate change. In other cases, we ought to continue to raise the profile of already well-known risks, like nuclear warfare, to make sure that the attention of the international community does not slip.

2.3.2 Proposed interventions

There are many paths towards raising attention in the international system, and some strategies will be more plausible for certain actors than others. Here, we offer some approaches which appear to be generally promising. There will be others, not considered here, and not all of those listed will be appropriate to every actor.

EXISTENTIAL RISK NEGLIGENCE AS A CRIME AGAINST HUMANITY

It might be possible to enshrine more binding commitments to reduce existential risk, but this would be difficult until the member states of the General Assembly become more supportive of decisive action on existential risk than they currently are. For example, it could in theory be made a crime against humanity to recklessly and negligently take action which creates a significant risk of extinction for humanity. At the moment, only actualised acts which have been committed against an identifiable group can be prosecuted, which would not include risks to humanity until it became too late.

Introducing such as crime would require an amendment to the Rome Statute establishing the International Criminal Court (ICC), which would be an arduous process. Moreover, the Rome Statute remains unratified by key member states. This means that even if negligently exposing the world to existential risks were prosecutable by the ICC, it would not necessarily change the behaviour of key risk takers. As a result, we do not currently recommend pursuit of an addition to the crimes against humanity as a vehicle for reducing existential risk, but this option indicates the sort of activity which might be possible with a sufficiently broad global mandate.

We consider three strategies which make use of existing infrastructures, and two which might involve the creation of alternative institutions within the UN System.

2.3.2.1 Statements or declarations

A number of statements or declarations concerning existential risks have already come from prominent sources: Lord Martin Rees, former President of the Royal Society, has repeatedly called attention to the area, as have Professor Stephen Hawking, Professor Nick Bostrom, and technologists Elon Musk and Bill Gates. Some limited declarations have also been made by UNESCO in its formal declaration concerning the responsibilities of present generations to future generations and by Secretary General Ban Ki-moon in informal remarks. Despite this, more that could be achieved through statements which boost the profile of work to reduce existential risk and which apply to existential risks explicitly and as a whole.

One option would be a discussion or an expression of concern in a major international forum. The United Nations General Assembly (UNGA) would probably be an appropriate venue for such a resolution, and as a representative body it would have the advantage that such discussions may be more likely to reach the attention of all member states. Alternatively, it is conceivable that the United Nations Security Council (UNSC) could be an appropriate venue for discussion, or for a statement by the President of the Security Council. However, it is debatable whether the UNSC's remit of the 'maintenance of peace and security' covers existential risk for humanity as a whole, or whether it is limited to specific security issues.

In the best case, such a resolution would highlight the possibility of human extinction, identify some possible areas of concern (while acknowledging unknown unknowns), and express the desire to take steps as an international community to reduce the size of the risk. Admittedly, it is unlikely that this would have a noticeable direct effect, since most member states would continue to reduce risk to the same degree as they had been intending to already. However, such a resolution might serve as a tool for use in future negotiations, to act as evidence of broad existing support for existential risk reduction, and as part of a process of educating a broader audience about existential risk. Another option might be to include slightly stronger wording calling on member states to act to reduce existential risk, which would likely have similar outcomes.

The process of negotiating a statement would lead to many policy-makers being introduced to and educated about existential risks. Significant international cooperation can occur through growing national awareness long before the establishment of UN-level agreements, potentially within the EU for example.

2.3.2.2 Reports

A number of bodies could commission or produce reports on existential risk, potentially building on existing work, including this report, and drawing on the outputs of ongoing research. Many issues remain to be explored in increasing detail, including the risk profiles and drivers of key risks, and the best strategies for reducing them. Actors who might work on these reports include academic institutions, specialist institutes like the Centre for the Study of Existential Risk at the University of Cambridge or the Future of Humanity Institute at the University of Oxford, national governments, strategic research centres, and parts of the UN System. Existing networks of interested parties, such as the Bulletin of the Atomic Scientists, the Munich Security Conference, or the Pugwash Conference,

might be well-suited to either creating or distributing such knowledge, in order to unlock the experience of their members.

Further reports on this topic offer several opportunities. First, they spread information about existential risks further through the international community. Reframing ideas for new or wider audiences can gradually raise the profile of these ideas. Second, each further exploration gives an opportunity for a new perspective on the issues and potentially significant improvements to the best identified strategies. Third, the process of developing recommendations can be extremely beneficial to the organisations drafting the reports, allowing them to develop their capabilities and knowledge, and encouraging them to probe adjacent organisations about their level of preparedness. Fourth, the fact that these ideas receive sustained and visible attention from serious thinkers will continue to increase attention to and support for the reduction of these risks.

However, it is important to be aware of the potential opportunity costs of commissioning reports. Too many reports, even those which set out to avoid doing so, fail to encourage action and end up gathering dust in cabinets.

2.3.2.3 Training courses

There are many venues that currently provide courses and educational opportunities to individuals with the power to take action on existential risks. These venues, such as providers of executive education on domestic and international security, could easily include a one-day course on existential risk as part of their classes. For example, the authors of this report recently taught a session at the Geneva Centre for Security Policy as part of its 8-week New Issues In Security course. Some bodies within the UN System, such as UNIDIR, may also be well-placed to train members of the international community on existential risk.

These courses will probably be most effective if they present the best known risk profiles of the largest existential risks (with appropriate uncertainty), communicate the importance of the area, include steps that can be taken by the international community to reduce the risks, and encourage participants to apply the things they have learned to their own jobs.

Developing and delivering training on existential risk is likely to be a fairly gradual process, and one might expect limited behavioural change as a result of the training, unless there is a clear way to tie its lessons into people's daily work routines. It may be useful to take advantage of digital distribution, potentially as a MOOC, to scale up the level of training that can be offered more quickly.

2.3.2.4 Political representation for Future Generations

Existential risk reduction is beneficial for those alive today, but an overwhelming amount of the value accrues to those who would live in the future. It might be that empowering actors to act as political representatives for the concerns of future generations would ensure that more weight was given to concerns about existential risks.

There have been a number of proposals for representation of future generations, at a number of levels of government.[47] At the level of the United Nations as a whole, the option of a High Commissioner for Future Generations was considered in the report "Intergenerational solidarity and the needs of future generations", produced by the Secretary General's office in 2013. This role would help bring attention to the needs of future generations and the steps that the current generation could take to safeguard them in a manner analogous to the High Commissioners for Human Rights and Refugees, although potentially in a smaller capacity. However, insofar as many of the effects of the current generation on future generations are mediated through either the environment or education and culture, the role might overlap with that of the UN Environment Programme or UNESCO. Moreover, attempts to create a High Commissioner appear to be currently stalled.

At a national level, other strategies are available for political representation of future generations. It is possible to assign ministerial positions representing future generations (as is done in Sweden), form a committee of parliamentarians responsible for the future (as in Finland), create a future generations commissioner (as in Wales), or embed a commitment to the rights of future generations in the constitution (as in Japan). In these cases, some variety of office is charged with the task of evaluating the impacts of other departments' activities on future generations and advocating for policies which respect the interests of the future. In many other countries, such as Namibia, Brazil, the Philippines, or Bhutan, there are similar provisions focused specifically on stewarding the environment for the sake of future generations. The environment is an important component of the interests and rights of future generations, but ideally a body would have a broader mandate to consider all sorts of choices we make now that have consequences for the future.

For all of these approaches, there is a significant risk. If the position supposedly representing future generations is politically weak, it is easy for other political actors to shrug off its warnings or advice. If it is poorly resourced, or lacks sufficient security clearance, it may lack the capacity to properly engage with all the activities of government. It is also plausible

that the presence of someone whose job it is to think about the future might make other departments feel that it is no longer necessary for them to bear these issues in mind themselves. This could, potentially, result in less attention being paid to future generations than before as an unintended consequence. Moreover, there are many interpretations of a commissioner for future generations which might place relatively little emphasis on existential risks relative to other, less neglected issues like preserving cultural heritage or reducing pollution.

Unfortunately, there is little evidence regarding the practical effectiveness of these institutions, and it would be difficult to gather such evidence systematically. Relatively few countries have experimented with such systems, and those that have are unlike other countries in many ways.

2.3.2.5 UN Office of Existential Risk Reduction

It may be possible to create a targeted institution within the UN System which concerns itself chiefly with existential risk.[48] Existential risk touches on the mandates of many different parts of the System, and as a result a central coordinating role could be very valuable in ensuring that balls do not get dropped between agencies. Such an office could support investigation of existential risks, connect ongoing research with the parts of the UN system which help coordinate between states engaged in risky activities, and swiftly call the attention of the UN System to existential risks if they arose unexpectedly. This might be critical in extreme scenarios.

It would be natural for a team working on existential risk to start out within an existing body, while it establishes itself and develops its capabilities. Several organisations within the UN system might be a natural home for such a team: for example, UNISDR's work on disaster risk might give it the right capabilities, as might OCHA's work on humanitarian response (although both naturally tend to focus on more salient and regularly-occurring risks). As the team becomes more self-sufficient, or as the magnitude of existential risks grows with the development of new technologies, it might be appropriate to spin it out to form its own office. Alternatively, it might make sense for the Secretary General to establish a High-Level Panel on existential risk.

This approach may have many of the same risks as that of establishing a commissioner – a weak institution may simply give others the sense that these risks are no longer their problem without having the power to achieve anything. There are likely to be significant barriers to coordination between organisations. For example, there may well be disagreements about jurisdiction over particular issues. Moreover, such an office would require sustained financing. In order to acquire such financing, it would need to demonstrate its value and also offer guarantees not to consume the resources of other existing institutions. It may also need to be championed by a member state which represents the interests of the broader international community.

2.3.3 Impact of the intervention

It is extremely difficult to model the impact which attention-building within the international community might have. On the one hand, it seems entirely possible that each of the options above could be successfully implemented and yet lead to no tangible results. Moreover, it seems conceivable that significant reductions in existential risk might happen without the UN getting involved. These steps therefore seem neither necessary nor sufficient for large risk reduction.

On the other hand, it is just as easy to imagine such efforts playing a decisive role, by lending legitimacy to the efforts of individuals working to reduce existential risk and positioning their work within a broader landscape. Moreover, if the process of building attention galvanises and empowers even one senior individual to make it their mission to reduce the risks of human extinction then the impact could be substantial. As a result, we recommend that individuals and organisations with particularly good opportunities or particularly suitable skills should make attention-building their priority. It is of course important for this process that those who focus on building attention remain in close communication with experts in the field.

It is worth noting that it is very difficult to 'take back' efforts to increase awareness, and that the framing of an approach can be highly significant. Therefore individuals or organisations who plan to invest significantly in drawing attention to existential risk would do well to engage a wide range of other interested parties before taking action. For example, we would caution against approaches which risk alienating the research and technology communities who will be essential in managing risks responsibly.

2.3.4 Ease of making progress

Building international attention on a global issue takes a long time, and one should expect the process to stall or appear to have stalled several times before it succeeds. This makes it difficult to know whether the current approach is working.

There is, however, a fairly good track record of bringing the attention of the international community to major risks and threats, albeit with large delays. This has happened with climate change, acid rain, ozone depletion, nuclear weapons and biological weapons (to name a few examples). Unfortunately, success in galvanising international action often ul-

INTERVENTIONS UNDER CONSIDERATION WHICH DID NOT REACH THE FINAL STAGE

The following interventions are a sample of those which reached later stages of development but were discarded. We have not done enough analysis on any of these to definitively decide whether they are likely to be a good or bad idea. You can see more in the appendix.

INTERVENTION	DESCRIPTION
Research funding.	Dramatically increase research resources targeting existential risk reduction.
Publishers' agreement on dual-use.	Build international agreement between publishers to respect decisions not to publish out of dual-use concerns.
Stockpiling agreements.	Create international agreements on levels of and distribution of stockpiles of essentials in catastrophic scenarios.
Tail-risk climate change treaty.	Create international agreements for decisive action in the event of extreme climate change.
Identifying recovery knowledge.	Catalogue the knowledge and capacities required for recovery (and store records).
Responsible whistleblowing support.	Create institutions to support and protect responsible whistleblowers.
Forecasting body.	Develop a UN or non-governmental forecasting body concerned with existential risk.
Meta-institutional red/blue team exercises.	Fund an exercise where a 'red team' tries to identify failure modes of institutions and 'blue team' tries to fix.

timately depends on a particularly traumatic shared event. For example, the horror of the Second World War may have created the conditions necessary for the Geneva Conventions' attention to civilians in war. In the case of existential risks, an event that made it clear how urgent existential risk reduction was might well come too late.

It can be easier to get agreement on more general principles, such as those underpinning the Paris Agreement. However, general principles are sometimes less effective than agreeing on specific actions.

It is often easier to start with smaller-scale proofs of concept within a regional community that is particularly able to coordinate, such as the European Union. However, the truly global nature of existential risks means that frameworks which are genuinely owned by the whole world and are deeply inclusive are likely to be more valuable in the long run.

2.3.5 What next steps can people take?

Individuals can take a number of easy steps. First, they can spread knowledge of the issues by discussing them with their colleagues or by passing on this or similar reports. Second, they can work out how to apply these ideas to their day-to-day work: for people who work as part of the international community, bearing in mind the reduction of existential risk might help them to make better marginal decisions during normal work, which add up in the long run.

The largest effects will probably come from individuals who hear about these ideas and arguments and decide to make responding to them their primary focus. The field of existential risk reduction is rapidly expanding as funding for work in the area grows. It would be very valuable for an expert or experts in the international community to establish an advocacy group, and it seems likely that private funding could be acquired by a sufficiently skilful team.

Endnotes – 2. Recommendations

1. See David R. Morrow, "Ethical Aspects of the Mitigation Obstruction Argument against Climate Engineering Research," Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 372, no. 2031 (December 28, 2014): 20140062, doi:10.1098/rsta.2014.0062.

2. J. G. Shepherd, Geoengineering the Climate: Science, Governance and Uncertainty (Royal Society, 2009); S. Schäfer et al., "The European Transdisciplinary Assessment of Climate Engineering (EuTRACE): Removing Greenhouse Gases from the Atmosphere and Reflecting Sunlight Away from Earth," Institute for Advanced Sustainability Studies, Potsdam, 2015; Bipartisan Policy Centre, "Task Force On Climate Remediation Research," 2011; National Academy of Sciences, Climate Intervention: Reflecting Sunlight to Cool Earth (Washington, D.C.: National Academies Press, 2015).

3. David W. Keith, Riley Duren, and Douglas G. MacMartin, "Field Experiments on Solar Geoengineering: Report of a Workshop Exploring a Representative Research Portfolio," Phil. Trans. R. Soc. A 372, no. 2031 (December 28, 2014): 20140175, doi:10.1098/rsta.2014.0175.

4. Andy Parker, "Governing Solar Geoengineering Research as It Leaves the Laboratory," Phil. Trans. R. Soc. A 372, no. 2031 (December 28, 2014): 20140173, doi:10.1098/rsta.2014.0173.

5. There appears to be expert disagreement about the extent to which international law does constrain SRM. See for example Edward A. Parson, "Climate Engineering in Global Climate Governance: Implications for Participation and Linkage," Transnational Environmental Law 3, no. 1 (April 2014): 89–110, doi:10.1017/S2047102513000496; and Chiara Armeni and Catherine Redgwell, "International Legal and Regulatory Issues of Climate Geoengineering Governance: Rethinking the Approach," 2015.

6. Parson, "Climate Engineering in Global Climate Governance."

7. For discussion see Steve Rayner et al., "The Oxford Principles," Climatic Change 121, no. 3 (January 24, 2013): sec. 8, doi:10.1007/s10584-012-0675-2.

8. Ibid., 508.

9. Parker, "Governing Solar Geoengineering Research as It Leaves the Laboratory."

10. Edward A. Parson and David W. Keith, "End the Deadlock on Governance of Geoengineering Research," Science 339, no. 6125 (March 15, 2013): 279, doi:10.1126/science.1232527.

11. Rayner et al., "The Oxford Principles."

12. Anna-Maria Hubert and David Reichwein, "An Exploration of a Code of Conduct for Responsible Scientific Research Involving Geoengineering" (Institute for Advanced Sustainability Studies, Potsdam; Institute of Science, Innovation, and Society, University of Oxford, 2015).

13. Rayner et al., "The Oxford Principles."

14. Edward A. Parson and David W. Keith, "End the Deadlock on Governance of Geoengineering Research," Science 339, no. 6125 (March 15, 2013): 1278–79, doi:10.1126/science.1232527.

15. Stefan Schäfer et al., "Field Tests of Solar Climate Engineering," Nature Climate Change 3, no. 9 (September 2013): 766–766, doi:10.1038/nclimate1987.

16. See http://www.openphilanthropy.org/research/cause-reports/geoengineering#Who_else_is_working_on_this

17. Donald A. Henderson, "The Looming Threat of Bioterrorism," Science 283, no. 5406 (February 26, 1999): 1279–82, doi:10.1126/science.283.5406.1279.

18. See for example Martin J. Rees, Our Final Century: Will Civilisation Survive the Twenty-First Century? (London: Arrow, 2004); Ali Nouri and Christopher F. Chyba, "Biotechnology and Biosecurity," in Global Catastrophic Risks, ed. Nick Bostrom and Milan M. Ćirković (Oxford University Press, 2008).

19. World Health Organisation, "WHO Activities in Avian Influenza and Pandemic Influenza Preparedness," December 2006.

20. Lawrence O. Gostin and Rebecca Katz, "The International Health Regulations: The Governing Framework for Global Health Security," The Milbank Quarterly 94, no. 2 (June 1, 2016): 264–313, doi:10.1111/1468-0009.12186.

21. World Health Organization, "Smallpox, Bioterrorism, and the World Health Organization," June 2006, http://www.who.int/global_health_histories/seminars/paper02.pdf.

22. World Health Organization, "Biorisk Management: Core Documents," 2016, http://www.who.int/ihr/publications/bioriskmanagement_1/en/.

23. See World Animal Health Organisation, "Biological Threat Reduction Strategy: Strengthening Global Biological Security," 2015, http://www.oie.int/fileadmin/Home/eng/Our_scientific_expertise/docs/pdf/EN_FINAL_Biothreat_Reduction_Strategy_OCT2015.pdf.

24. See for example World Organisation for Animal Health, "Terrestrial Animal Health Code," 2016, chap. 5.8.

25. Food and Agriculture Organization, "The Global Strategy for Prevention and Control of H5N1 Highly Pathogenic Avian Influenza," 2008.

26. World Health Organisation, "Considerations on Exercises to Validate Pandemic Preparedness Plans," n.d.

27. Ibid.

28. See for example World Health Organisation, "WHO Activities in Avian Influenza and Pandemic Influenza Preparedness"; OiE, "Disease Introduction Simulation Exercises," n.d., http://www.oie.int/animal-health-in-the-world/the-world-animal-health-information-system/simulation-exercises/2016/; Food and Agriculture Organization, "FAO/WHO/USAID Development of Simulation Exercises on Avian Influenza in Animal and Human Populations in Europe and Eurasia," n.d.

29. World Organisation for Animal Health, "Biological Threat Reduction Strategy: Strengthening Global Biological Security"; Interpol, "Bioterrorism," n.d., http://www.interpol.int/Crime-areas/CBRNE/Bioterrorism; World Health Organization, "Public Health Response to Biological and Chemical Weapons: WHO Guidance (2004)".

30. See for example Interpol, "Bioterrorism International Tabletop Exercise: Black Death," 2009; World Health Organization, "Smallpox, Bioterrorism, and the World Health Organization."

31. World Organisation for Animal Health, "Biological Threat Reduction Strategy: Strengthening Global Biological Security."

32. Wilton Park, "The 2014-2015 Ebola Outbreak: Lessons for Response to a Deliberate Event (WP1496)," https://www.wiltonpark.org.uk/event/wp1496/.

33. World Health Organization, "Smallpox, Bioterrorism, and the World Health Organization."

34. Rebecca Katz et al., "Global Health Security Agenda and the International Health Regulations: Moving Forward," Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science 12, no. 5 (September 1, 2014): 231–38, doi:10.1089/bsp.2014.0038.

35. Lawrence O. Gostin et al., "Toward a Common Secure Future: Four Global Commissions in the Wake of Ebola," PLOS Med 13, no. 5 (May 19, 2016): e1002042, doi:10.1371/journal.pmed.1002042.

36. The United Nations Office at Geneva, "Intersessional Work," 2016, http://www.unog.ch/80256EE600585943/(httpPages)/1B69CE1F0B030DA0C1257F39003E9590?OpenDocument.

37. David P. Fidler and Lawrence O. Gostin, "The New International Health Regulations: An Historic Development for International Law and Public Health," The Journal of Law, Medicine & Ethics 34, no. 1 (March 1, 2006): 85–94, doi:10.1111/j.1748-720X.2006.00011.x.

38. Gostin and Katz, "The International Health Regulations."

39. For an overview see World Bank, "Pandemic Emergency Facility: Frequently Asked Questions," Text/HTML, World Bank, accessed June 22,

2016, http://www.worldbank.org/en/topic/pandemics/brief/pandemic-emergency-facility-frequently-asked-questions.

40. See WHO, "WHO | Ebola Virus Disease Outbreak," WHO, accessed June 21, 2016, http://www.who.int/csr/disease/ebola/en/.

41. World Bank, "World Bank Group Launches Groundbreaking Financing Facility to Protect Poorest Countries against Pandemics," Text/HTML, World Bank, accessed June 21, 2016, http://www.worldbank.org/en/news/press-release/2016/05/21/world-bank-group-launches-groundbreaking-financing-facility-to-protect-poorest-countries-against-pandemics.

42. Lawrence O. Gostin and Rebecca Katz, "The International Health Regulations: The Governing Framework for Global Health Security," The Milbank Quarterly 94, no. 2 (June 1, 2016): 264–313, doi:10.1111/1468-0009.12186.

43. World Bank, 'Pandemic Emergency Facility: Frequently Asked Questions' World Bank <http://www.worldbank.org/en/topic/pandemics/brief/pandemic-emergency-facility-frequently-asked-questions> [accessed 22 June 2016].

44. Olga Jones B., "Pandemic Risk" (World Bank, 2014), 17.

45. Lawrence O. Gostin and Rebecca Katz, "The International Health Regulations: The Governing Framework for Global Health Security," The Milbank Quarterly 94, no. 2 (June 1, 2016): 264–313, doi:10.1111/1468-0009.12186.

46. United Nations Framework Convention on Climate Change.

47. For example, see Kavka, Gregory, and Virginia Warren. "Political representation for future generations." Environmental Philosophy 21 (1983): 28.; Ekeli, Kristian Skagen. "Giving a voice to posterity–deliberative democracy and representation of future people." Journal of Agricultural and Environmental Ethics 18.5 (2005): 429-450.; Tonn, B., 1996, "A Design for Future-Oriented Government," Futures, Vol. 28, No. 5, pp. 413-431.; Wolfe, M. W. (2008). The shadows of future generations. Duke Law Journal, 1897-1932.

48. Richard Posner discusses a somewhat related proposal, though more limited in scope in his book Posner, Richard A. Catastrophe: risk and response. Oxford University Press, 2004.

Appendix - Methodology

This appendix outlines our approach to the process of identifying strategies for risk reduction, and suggests blind-spots we expect other may later be able to fill.

A systematic approach to existential risk reduction needs to identify the strategies which are collectively likely to offer the largest reduction in risk for a given cost, in terms of time, money, and political will. However, this is difficult. The processes governing existential risks are complicated and unfold over years and decades. Similarly, political processes confound even experienced forecasters. We should therefore, as a general principle, adopt a cautious approach which reflects the deep uncertainty of our position.

Our approach was first to generate a large number of unfiltered potential recommendations and then gradually to prioritise and develop them. In both the generation phase and the prioritisation phase, we consulted a variety of experts and we used classifications of strategies to cross-check that we did not miss large areas of ideas. We expect that, despite our efforts, there will remain areas that are valuable but have not been explored.

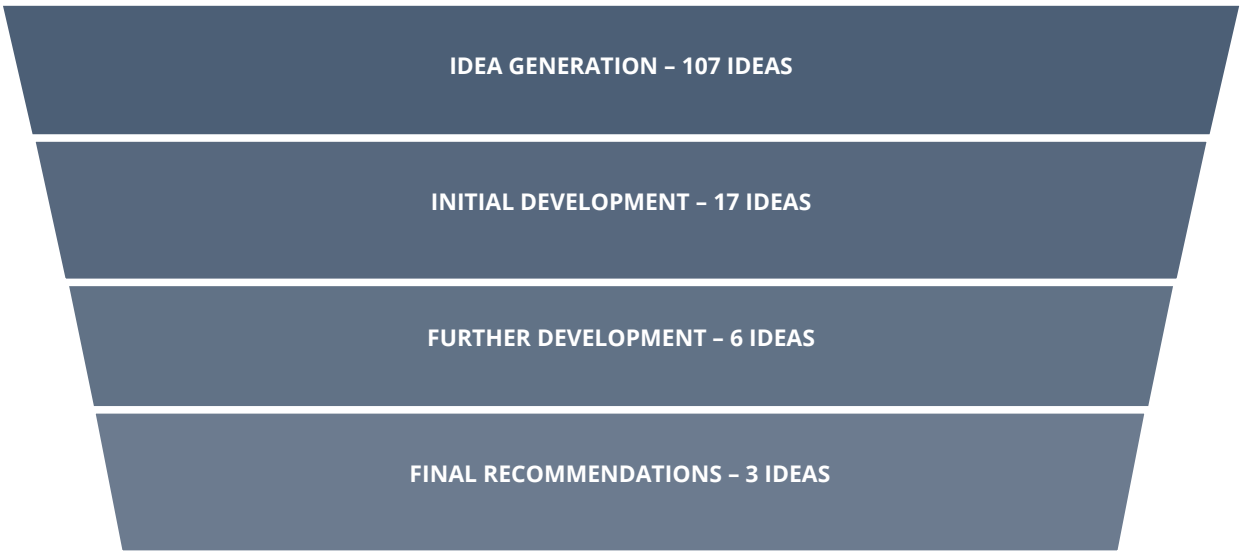
PHASE 1: IDEA GENERATION – 107 PROPOSALS

With the three main classifications identified in the first section in mind, we asked a range of experts to offer ideas for how best to reduce existential risk. At this stage, the participants were asked to focus on creativity rather than reliability in their suggestions. There were three main sources of suggestions: internal brainstorming by the authors, a workshop hosted in February 2016 at the Oxford Martin School, and a joint working-session in June 2016 of the Future of Humanity Institute at the University of Oxford and the Centre for the Study of Existential Risk at the University of Cambridge.

Afterwards, we filtered the list of suggestions to identify the most promising interventions for further study. We excluded interventions which:

- Were duplicates or similar to other proposals.
- Were too general to imply a clear course of action.
- Were too specific, such that they could only really be pursued by one particular actor.
- Were not relevant to the international community (for example, if they were mostly relevant to a specific research community).

The team then independently scored each intervention, presented in random order to each team member, considering the difficulty of implementation, the amount of value it would create if implemented, and the size of the resources already going towards it.



At this point, the top 15 interventions based on aggregated score were selected for further analysis. The team selected two additional interventions which had not quite made the cut but were representatives of categories which were otherwise no longer in the set of ideas to be developed further.

PHASE 2: INITIAL DEVELOPMENT — 17 PROPOSALS

After identifying the top interventions, we investigated each in detail, assessing:
The main mechanism of the intervention.

- The probable size of the effect.
- The variants to the intervention that might be used.
- The main institutions the intervention interacts with.
- The other projects or work most similar to the intervention.
- The timescales over which the intervention might happen.
- The path towards implementing the intervention.
- The tractability of the intervention.

Based on this analysis, we discussed the most promising interventions and selected six ideas that were suitable for further investigation.

INITIAL INTERVENTION	DESCRIPTION	DECISION
Training in existential risks.	Run a short course at, e.g., UNIDIR, to familiarise diplomats with existential risk issues.	Make a sub-point of building international attention to existential risk.
Liability insurance in gain-of-function research of concern.	Require research creating potentially risky pathogens to cover the expected cost to society in the research budget.	Drop. Plausible idea, but a topic for synthetic biology community specifically, and the recent NSABB process has closed.
Incorporate future generations in cost-benefit analysis.	When making government cost-benefit decisions, ensure that long-run future effects are represented.	Stress the political institutions aspect rather than technical process. Make a sub-point of building international attention to existential risk.
Stockpiling agreements.	Create international agreements on levels of and distribution of stockpiles of essentials in catastrophic scenarios.	Drop. Ensuring stockpiles are fairly distributed internationally in case of global catastrophe may be difficult. Unclear if large stockpiles are cost-effective.
Engineered pandemic planning.	Improve specific planning for pandemic scenarios resulting from engineered pathogens.	Advance to the next stage.
Tail-risk climate change treaty.	Create international agreements for decisive action in the event of extreme climate change.	Advance to the next stage, but focus on governance for geoengineering.
Research funding.	Dramatically increase research resources targeting existential risk reduction.	Advance to the next stage.
Identifying recovery knowledge.	Catalogue the knowledge and capacities required for recovery (and store records).	Drop. More appropriate as a standalone project. Some overlap with existing work.
Responsible whistleblowing support.	Create institutions to support and protect responsible whistleblowers.	Drop. Difficult to do responsibly in a way that builds on the existing system.
Forecasting body.	Develop a UN or non-governmental forecasting body concerned with existential risk.	Drop. Plausibly dominated by existential risk organisation which can choose how much to invest in forecasting relative to other things.
Lobby the UN.	Lobby the UN on existential risk. Many organs may be relevant.	Make a sub-point of building international attention to existential risk.
Pre-publication dual-use scanning.	Scan papers pre-publication for dual-use content to reduce risk from publication.	Advance to the next stage. Focus on aspect of coordination problem between publishers who might think someone else will publish if they do not.

INITIAL INTERVENTION (CONT.)	DESCRIPTION (CONT.)	DECISION (CONT.)
X-prize for pandemic surveillance.	Create an x-prize for, e.g., low-cost DNA sequencer for virus surveillance.	Drop. Existing funding is sufficient that marginal resources might have only a small effect.
Meta-institutional red/blue team exercises.	Fund an exercise where a 'red team' tries to identify failure modes of institutions and 'blue team' tries to fix.	Drop. Difficult to do responsibly given confidential and classified information.
UN existential risk centre.	Fund or establish a UN centre tasked with managing existential risk.	Make a sub-point of building international attention to existential risk.
Impact of technology on future of diplomacy.	Research the impact of technologies (e.g., blockchain, lie detection) on the future of diplomacy.	Drop. Make a component of research funding intervention.
Engineered pandemic planning.	Improve specific planning for pandemic scenarios resulting from engineered pathogens.	Fully develop.
World Bank Pandemic Emergency Facility.	Endorse, fund, and promote the World Bank's Pandemic Emergency Facility.	Develop as side-note, acknowledging flaws with the proposal.
Geoengineering governance.	Establish international norms for appropriate geoengineering as the technologies develop.	Fully develop.
Research funding.	Dramatically increase research resources targeting existential risk reduction.	Fully develop. Dropped after workshop feedback.
International attention building.	Increase attention to and knowledge about existential risks in international venues.	Fully develop.
Publishers' agreement on dual-use.	Build international agreement between publishers to respect decisions not to publish out of dual-use concerns.	Drop. Export controls make international aspect possibly unnecessary. Improvements in pre-funding dual-use screening will reduce need.

PHASE 3: FURTHER DEVELOPMENT – 6 PROPOSALS

At this point, a deeper analysis was carried out for six interventions. For each, we considered in further detail their likely impact, their tractability, and existing programmes carrying out similar work. We also engaged subject-specialists in each area.

This allowed us to identify the final recommendations, which were developed with heavy input from subject-specific experts. These recommendations can be found in Section 2.

DEVELOPED INTERVENTION	DESCRIPTION	DECISION
Engineered pandemic planning.	Improve specific planning for pandemic scenarios resulting from engineered pathogens.	Fully develop.
World Bank Pandemic Emergency Facility.	Endorse, fund, and promote the World Bank's Pandemic Emergency Facility.	Develop as side-note, acknowledging flaws with the proposal.
Geoengineering governance.	Establish international norms for appropriate geoengineering as the technologies develop.	Fully develop.
Research funding.	Dramatically increase research resources targeting existential risk reduction.	Fully develop. Dropped after workshop feedback.
International attention building.	Increase attention to and knowledge about existential risks in international venues.	Fully develop.
Publishers' agreement on dual-use.	Build international agreement between publishers to respect decisions not to publish out of dual-use concerns.	Drop. Export controls make international aspect possibly unnecessary. Improvements in pre-funding dual-use screening will reduce need.

HAVE WE IDENTIFIED THE BEST INTERVENTIONS?

The range of possible interventions is very large, and the literature on existential risks is not well developed. We have therefore not been comprehensive in surveying possibilities. Although our brainstorming processes did begin to create some overlapping ideas, it is very likely that other groups attacking the issue from a different perspective might have generated a different set of ideas. There may therefore be excellent interventions that have not been included here.

Moreover, a full assessment of these proposals may require deep domain expertise. We have made significant efforts to engage experts in specific fields relevant to the interventions, but it may be that the only way to get a full picture would be to have a single expert knowledgeable about both existential risk and specific intervention areas. Errors may have entered our process as a result.

Nonetheless, we believe that this process has highlighted opportunities that are more promising than we might have otherwise expected, and which are worth addressing swiftly.

Existential risks – those that could curtail humanity's long-term potential – are some of the most serious geopolitical challenges in the 21st century.

From nuclear war and the potential devastation of a nuclear winter, to the risks of accidents with emerging technologies, the legacy we leave future generations cannot be taken for granted.

This report identifies three important steps that can be taken to reduce existential risks today. It outlines the main considerations behind each proposal and identifies strategies for moving forwards. The recommendations are:

- Develop governance for geoengineering research.
- Establish scenario plans and exercises for severe engineered pandemics at the international level.
- Build international attention to and support for existential risk reduction.