



The case for reducing existential risks

By Benjamin Todd · Published October 2017 · Last updated June 2022

The recording may not reflect the most recent changes to this article.

In 1939, Einstein wrote to Roosevelt:¹

It may be possible to set up a nuclear chain reaction in a large mass of uranium...and it is conceivable — though much less certain — that extremely powerful bombs of a new type may thus be constructed.

Just a few years later, these bombs were created. In little more than a decade, enough had been produced that, for the first time in history, a handful of decision-makers could destroy civilisation.

Humanity had entered a new age, in which we faced not only existential risks² from our natural environment, but also the possibility that we might be able to extinguish ourselves.



Prefer a podcast?

Since publishing this article, we recorded two **podcast episodes** with Dr Toby Ord, an Oxford philosopher and trustee of 80,000 Hours, about existential threats. We think they are at least as good introductions as this article — maybe better. Listen to them here:

- [Toby Ord on the precipice and humanity's potential futures](#)
- [Why the long-term future of humanity matters more than anything else, and what we should do about it](#)

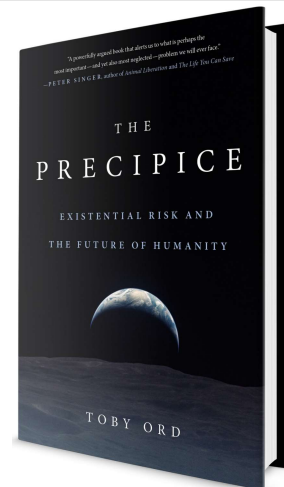
Prefer a book?

Dr Toby Ord, an Oxford philosopher and 80,000 Hours trustee, has recently published [*The Precipice: Existential Risk and the Future of Humanity*](#) which gives an overview of the moral importance of future generations, and what we can do to help them today.

We'll mail you the book (for free)

Join the 80,000 Hours newsletter and we'll send you a free copy of the book.

We'll also send you updates on our latest research, opportunities to work on existential threats, and news from the author.



If you're already on our newsletter, email us at book.giveaway@80000hours.org to get a copy.

In this new age, what should be our biggest priority as a civilisation? Improving technology? Helping the poor? Changing the political system?

Here's a suggestion that's not so often discussed: our first priority should be to *survive*.

So long as civilisation continues to exist, we'll have the chance to solve all our other problems, and have a far better future. But if we go extinct, that's it.

Why isn't this priority more discussed? Here's one reason: many people don't yet appreciate the change in situation, and so don't think our future is at risk.

Social science researcher [Spencer Greenberg](#) surveyed Americans on their estimate of the chances of human extinction within 50 years. The results found that many think the chances are extremely low, with over 30% guessing they're under 1 in 10 million.³

We used to think the risks were extremely low as well, but when we looked into it, we changed our minds. As we'll see, researchers who study these issues think the risks are over 1,000 times higher, and are probably increasing.

These concerns have started a new movement working to safeguard civilisation, which has been joined by Stephen Hawking, Max Tegmark, and new institutes founded by researchers at [Cambridge](#), [MIT](#), [Oxford](#), and elsewhere.

In the rest of this article, we cover the greatest risks to civilisation, including some that might be bigger than nuclear war and climate change. We then make the case that reducing these risks could be the most important thing you do with your life, and explain exactly what you can do to help. If you would like to use your career to work on these issues, we can also give one-on-one support.

Reading time: 25 minutes

How likely are you to be killed by an asteroid? An overview of naturally occurring existential threats

A 1-in-10-million chance of extinction in the next 50 years — what many people think the risk is — must be an underestimate. Naturally occurring existential threats can be estimated pretty accurately from history, and are much higher.

If Earth was hit by a kilometre-wide asteroid, there's a chance that civilisation would be destroyed. By looking at the historical record, and tracking the objects in the sky, astronomers can estimate the risk of an asteroid this size hitting Earth as about 1 in 5,000 per century.⁴ That's higher than most people's chances of being in a plane crash (about 1 in 5 million per flight), and already about 1,000 times higher than the 1-in-10-million risk that some people estimated.⁵

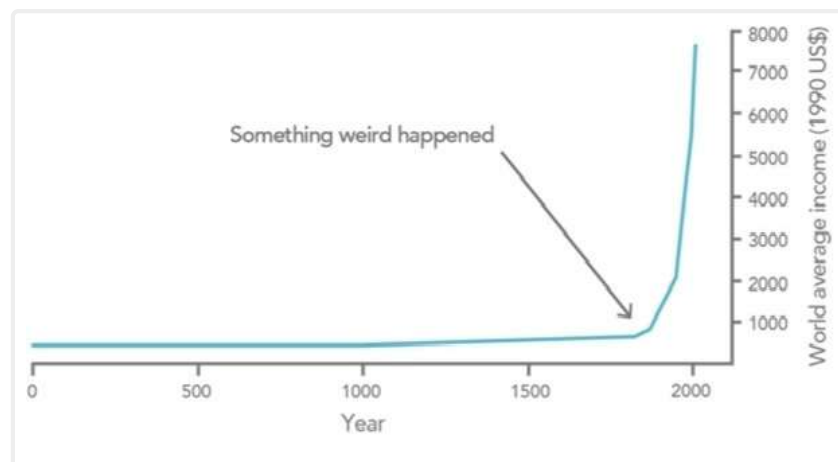
Some argue that although a kilometre-sized object would be a disaster, it wouldn't be enough to cause extinction, so this is a high estimate of the risk. But on the other hand, there are other naturally occurring risks, such as supervolcanoes.⁶

All this said, natural risks are still quite small in absolute terms. An upcoming paper by Dr Toby Ord estimated that if we sum all the natural risks together, they're very unlikely to add up to more than a 1 in 300 chance of extinction per century.⁷

Unfortunately, as we'll now show, the natural risks are dwarfed by the human-caused ones. And this is why the risk of extinction has become an especially urgent issue.

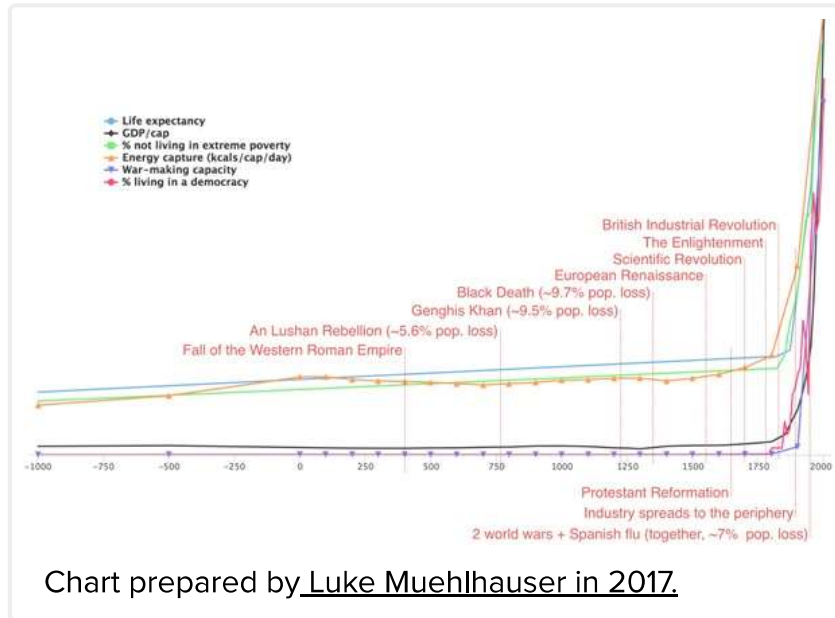
A history of progress, leading to the start of the most dangerous epoch in human history

If you look at history over millennia, the basic message is that for a long time almost everyone was poor, and then in the 18th century, that changed.⁸

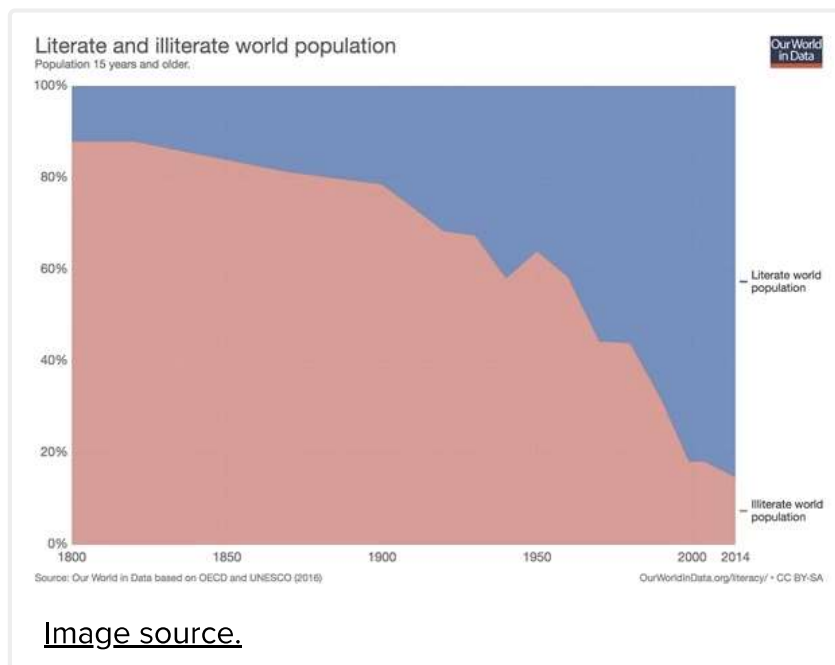


This was caused by the Industrial Revolution — perhaps the most important event in history.

It wasn't just wealth that grew. The following chart shows that over the long term, life expectancy, energy use and democracy have all grown rapidly, while the percentage living in poverty has dramatically decreased.⁹



Literacy and education levels have also dramatically increased:



People also seem to become happier as they get wealthier.

In *The Better Angels of Our Nature*, Steven Pinker argues that violence is going down. (Although our recent podcast with Bear Braumoeller — released after this article was published — looks at some reasons this might not be the case.)

Individual freedom has increased, while racism, sexism and homophobia have decreased.

Many people think the world is getting worse,¹⁰ and it's true that modern civilisation does some terrible things, such as factory farming. But as you can see in the data, many important measures of progress have improved dramatically.



More to the point, no matter what you think has happened in the past, if we look forward, improving technology, political organisation and freedom gives our descendants the potential to solve our current problems, and have vastly better lives.¹¹ It is possible to end poverty, prevent climate change, alleviate suffering, and more.

But also notice the purple line on the second chart: *war-making capacity*. It's based on estimates of global military power by the historian Ian Morris, and it has also increased dramatically.

Here's the issue: improving technology holds the possibility of enormous gains, but also enormous risks.

Each time we discover a new technology, most of the time it yields huge benefits. But there's also a chance we discover a technology with more destructive power than we have the ability to wisely use.

And so, although the present generation lives in the most prosperous period in human history, it's plausibly also the most dangerous.

The first destructive technology of this kind was nuclear weapons.

Nuclear weapons: a history of near misses

Today we all have North Korea's nuclear programme on our minds, but current events are just one chapter in a long saga of near misses.

We came close to nuclear war several times during the Cuban Missile Crisis alone.¹² In one incident, the Americans resolved that if one of their spy planes were shot down, they would immediately invade Cuba without a further War Council meeting. The next day, a spy plane was shot down. JFK called the council anyway, and decided against invading.

An invasion of Cuba might well have triggered nuclear war; it later emerged that Castro was in favour of nuclear retaliation even if "it would've led to the complete annihilation of Cuba." Some of the launch commanders in Cuba also had independent authority to target American forces with tactical nuclear weapons in the event of an invasion.

In another incident, a Russian nuclear submarine was trying to smuggle materials into Cuba when they were discovered by the American fleet. The fleet began to drop dummy depth charges to force the submarine to surface. The Russian captain thought they were real depth charges and that, while out of radio communication, the third world war had started. He ordered a nuclear strike on the American fleet with one of their nuclear torpedoes.

Fortunately, he needed the approval of other senior officers. One, Vasili Arkhipov, disagreed, preventing war.



Thank you Vasili Arkhipov.

Putting all these events together, JFK later estimated that the chances of nuclear war were “between one in three and even.”¹³

There have been plenty of other close calls with Russia, even after the Cold War, as listed on [this nice Wikipedia page](#). And those are just the ones we know about.

Nuclear experts today are just as concerned about tensions between India and Pakistan, which both possess nuclear weapons, as North Korea.¹⁴

The key problem is that several countries maintain large nuclear arsenals that are ready to be deployed in minutes. This means that a false alarm or accident can rapidly escalate into a full-blown nuclear war, especially in times of tense foreign relations.

Would a nuclear war end civilisation? It was initially thought that a nuclear blast might be so hot that it would ignite the atmosphere and make the Earth uninhabitable. Scientists estimated this was sufficiently unlikely that the weapons could be “safely” tested, and we now know this won’t happen.

In the 1980s, the concern was that ash from burning buildings would plunge the Earth into a long-term winter that would make it impossible to grow crops for decades.¹⁵ Modern climate models suggest that a nuclear winter severe enough to kill everyone is very unlikely, though it’s hard to be confident due to [model uncertainty](#).¹⁶

Even a “mild” nuclear winter, however, could still cause mass starvation.¹⁷ For this and other reasons, a nuclear war would be extremely destabilising, and it’s unclear whether civilisation could recover.

How likely is it that a nuclear war could permanently end human civilisation? It’s very hard to estimate, but it seems hard to conclude that the chance of a civilisation-ending nuclear war in the next century isn’t over 0.3%. That would mean the risks from nuclear weapons are greater than all the natural risks put together. ([Read more about nuclear risks.](#))



This is why the 1950s marked the start of a new age for humanity. For the first time in history, it became possible for a small number of decision-makers to wreak havoc on the whole world. We now pose the greatest threat to our own survival — that makes today the most dangerous point in human history.

And nuclear weapons aren't the only way we could end civilisation.

How big is the risk posed by climate change?

In 2015, President Obama said in his State of the Union address that “No challenge poses a greater threat to future generations than climate change.”

Climate change is certainly a major risk to civilisation.

The most likely outcome is 2–4 degrees of warming,¹⁸ which would be bad, but survivable for our species.

However, some estimates give a 10% chance of warming over 6 degrees, and perhaps a 1% chance of warming of 9 degrees.

So, it seems like the chance of a massive climate disaster created by CO₂ is perhaps similar to the chance of a nuclear war.

But as we argue in our problem profile on climate change, it looks unlikely that even 13 degrees of warming would directly cause the extinction of humanity. As a result, researchers who study these issues think nuclear war seems more likely to result in outright extinction, due to the possibility of nuclear winter, which is why we think nuclear weapons pose an even greater risk than climate change.

That said, climate change is certainly a major problem, and its destabilising effects could exacerbate other risks (including risks of nuclear conflict). This should raise our estimate of the risks even higher.

What new technologies might be as dangerous as nuclear weapons?

The invention of nuclear weapons led to the anti-nuclear movement just a couple decades later in the 1960s, and the environmentalist movement soon adopted the cause of fighting climate change.

What's less appreciated is that new technologies will present further catastrophic risks. This is why we need a movement that is concerned with safeguarding civilisation in general.

Predicting the future of technology is difficult, but because we only have one civilisation, we need to try our best. Here are some candidates for the next technology that's as dangerous as nuclear weapons.

Engineered pandemics

In 1918-1919, over 3% of the world's population died of the Spanish Flu.¹⁹ If such a pandemic arose today, it might be even harder to contain due to rapid global transport.

What's more concerning, though, is that it may soon be possible to genetically engineer a virus that's as contagious as the Spanish Flu, but also deadlier, and which could spread for years undetected.

That would be a weapon with the destructive power of nuclear weapons, but far harder to prevent from being used. Nuclear weapons require huge factories and rare materials to make, which makes them relatively easy to control. Designer viruses might be possible to create in a lab with a couple of people with biology PhDs. In fact, in 2006, The Guardian was able to receive segments of the extinct smallpox virus by mail order.²⁰ Some terrorist groups have expressed interest in using indiscriminate weapons like these. ([Read more about pandemic risks.](#))



Who ordered the smallpox? Credit: [The Guardian](#)

Artificial intelligence

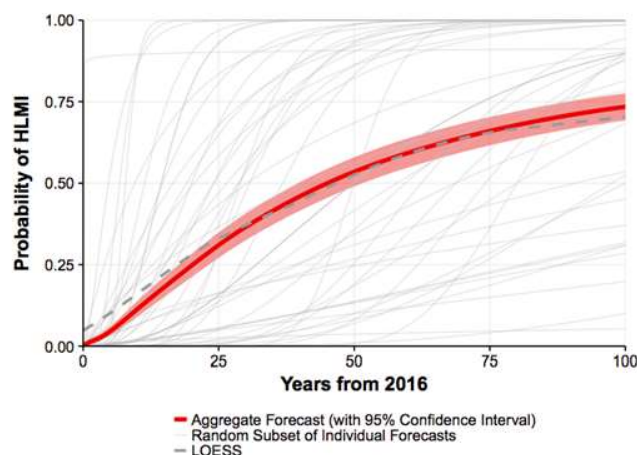
Another new technology with huge potential power is artificial intelligence.

The reason that humans are in charge and not chimps is purely a matter of intelligence. Our large and powerful brains give us incredible control of the world, despite the fact that we are so much physically weaker than chimpanzees.

So then what would happen if one day we created something much more intelligent than ourselves?

In 2017, 350 researchers who had published peer-reviewed research into artificial intelligence at top conferences were polled about when they believe that we will develop computers with human-level intelligence: that is, a machine that is capable of carrying out all work tasks better than humans.

The median estimate was that there is a 50% chance we will develop high-level machine intelligence in 45 years, and 75% by the end of the century.²¹





These probabilities are hard to estimate, and the researchers gave very different figures depending on precisely how you ask the question.²² Nevertheless, it seems there is at least a reasonable chance that some kind of transformative machine intelligence is invented in the next century. Moreover, greater uncertainty means that it might come sooner than people think rather than later.

What risks might this development pose? The original pioneers in computing, like Alan Turing and Marvin Minsky, raised concerns about the risks of powerful computer systems,²³ and these risks are still around today. We're not talking about computers "turning evil." Rather, one concern is that a powerful AI system could be used by one group to gain control of the world, or otherwise be misused. If the USSR had developed nuclear weapons 10 years before the USA, the USSR might have become the dominant global power. Powerful computer technology might pose similar risks.

Another concern is that deploying the system could have unintended consequences, since it would be difficult to predict what something smarter than us would do. A sufficiently powerful system might also be difficult to control, and so be hard to reverse once implemented. These concerns have been documented by Oxford Professor Nick Bostrom in [*Superintelligence*](#) and by AI pioneer [*Stuart Russell*](#).

Most experts think that better AI will be a hugely positive development, but they also agree there are risks. In the survey we just mentioned, AI experts estimated that the development of high-level machine intelligence has a 10% chance of a "bad outcome" and a 5% chance of an "extremely bad" outcome, such as human extinction.²¹ And we should probably expect this group to be positively biased, since, after all, they make their living from the technology.

Putting the estimates together, if there's a 75% chance that high-level machine intelligence is developed in the next century, then this means that the chance of a major AI disaster is 5% of 75%, which is about 4%. ([Read more about risks from artificial intelligence.](#))

Other risks from emerging technologies

People have raised concern about other new technologies, such as other forms of geo-engineering and [atomic manufacturing](#), but they seem significantly less imminent, so are widely seen as less dangerous than the other technologies we've covered. You can see a longer list of existential risks [here](#).

What's probably more concerning is the risks we haven't thought of yet. If you had asked people in 1900 what the greatest risks to civilisation were, they probably wouldn't have suggested nuclear weapons, genetic engineering or artificial intelligence, since none of these were yet invented. It's possible we're in the same situation looking forward to the next century. Future "unknown unknowns" might pose a greater risk than the risks we know today.

Each time we discover a new technology, it's a little like betting against a single number on a roulette wheel. Most of the time we win, and the technology is overall good. But each time there's also a small chance the technology gives us more destructive power than we can handle, and we lose everything.



Each new technology we develop has both unprecedented potential and perils. [Image source.](#)

What's the total risk of human extinction if we add everything together?

Many experts who study these issues estimate that the total chance of human extinction in the next century is between 1 and 20%.

In our [podcast episode with Will MacAskill](#) we discuss why he puts the risk of extinction this century at around 1%.

And in his 2020 book *[The Precipice: Existential Risk and the Future of Humanity](#)*, Toby Ord gives his guess at our total existential risk this century as 1 in 6 (or about 17%) — a roll of the dice. (Note, though, that Ord's definition of an existential catastrophe isn't exactly equivalent to human extinction; it would include, for instance, a global catastrophe that leaves the species unable to ever truly recover, even if some humans are still alive.) [Listen to our episode with Toby.](#)

His book provides the following table laying out his (very rough) estimates of existential risk from what he believes are the top threats:



Existential catastrophe via	Chance within next 100 years
Asteroid or comet impact	~ 1 in 1,000,000
Supervolcanic eruption	~ 1 in 10,000
Stellar explosion	~ 1 in 1,000,000,000
Total natural risk	~ 1 in 10,000
Nuclear war	~ 1 in 1,000
Climate change	~ 1 in 1,000
Other environmental damage	~ 1 in 1,000
'Naturally' arising pandemics	~ 1 in 10,000
Engineered pandemics	~ 1 in 30
Unaligned artificial intelligence	~ 1 in 10
Unforeseen anthropogenic risks	~ 1 in 30
Other anthropogenic risks	~ 1 in 50
Total anthropogenic risk	~ 1 in 6
Total existential risk	~ 1 in 6

These figures are about one million times higher than what people normally think.

What should we make of these estimates? Presumably, researchers interested in this topic only work on these issues because they think they're so important, so we should expect their estimates to be high (due to selection bias). While the figures are up for debate, we think a range of views, including those of MacAskill and Ord, is plausible.

Why helping to safeguard the future could be the most important thing you can do with your life

How much should we prioritise working to reduce these risks compared to other issues, like global poverty, ending cancer or political change?



At 80,000 Hours, we do research to help people find careers with positive social impact. As part of this, we try to find the most urgent problems in the world to work on. We evaluate different global problems using our problem framework, which compares problems in terms of:

- Scale — how many are affected by the problem
- Neglectedness — how many people are working on it already
- Solvability — how easy it is to make progress

If you apply this framework, we think that safeguarding the future comes out as the world's biggest priority. And so, if you want to have a big positive impact with your career, this is the top area to focus on.

In the next few sections, we'll evaluate this issue on scale, neglectedness and solvability, drawing heavily on Existential Risk Prevention as a Global Priority by Nick Bostrom and unpublished work by Toby Ord, as well as our own research.

First, let's start with the scale of the issue. We've argued there's likely over a 3% chance of extinction in the next century. How big an issue is this?

One figure we can look at is how many people might die in such a catastrophe. The population of the Earth in the middle of the century will be about 10 billion, so a 3% chance of everyone dying means the expected number of deaths is about 300 million. This is probably more deaths than we can expect over the next century due to the diseases of poverty, like malaria.²⁴

Many of the risks we've covered could also cause a "medium" catastrophe rather than one that ends civilisation, and this is presumably significantly more likely. The survey we covered earlier suggested over a 10% chance of a catastrophe that kills over 1 billion people in the next century, which would be at least another 100 million deaths in expectation, along with far more suffering among those who survive.

So, even if we only focus on the impact on the present generation, these catastrophic risks are one of the most serious issues facing humanity.

But this is a huge underestimate of the scale of the problem, because if civilisation ends, then we give up our entire future too.

Most people want to leave a better world for their grandchildren, and most also think we should have some concern for future generations more broadly. There could be many more people having great lives in the future than there are people alive today, and we should have some concern for their interests. There's a possibility that human civilisation could last for millions of years, so when we consider the impact of the risks on future generations, the stakes are millions of times higher — for good or evil. As Carl Sagan wrote on the costs of nuclear war in *Foreign Affairs*:

A nuclear war imperils all of our descendants, for as long as there will be humans. Even if the population remains static, with an average lifetime of the order of 100 years, over a typical time period for the biological evolution of a successful species (roughly 10 million years), we are talking about some 500 trillion people yet to come. By this criterion, the stakes are one million times greater for extinction than for the more modest nuclear wars that kill "only" hundreds of millions of people. There are many other possible measures of the potential loss—including culture and science, the evolutionary history of the planet, and the significance of the lives of all of our ancestors who contributed to the future of their descendants. Extinction is the undoing of the human enterprise.

We're glad the Romans didn't let humanity go extinct, since it means that all of modern civilisation has been able to exist. We think we owe a similar responsibility to the people who will come after us, assuming (as we believe) that they are likely to lead fulfilling lives. It would



be reckless and unjust to endanger their existence just to make ourselves better off in the short-term.

It's not just that there might be more people in the future. As Sagan also pointed out, no matter what you think is of value, there is potentially a lot more of it in the future. Future civilisation could create a world without need or want, and make mindblowing intellectual and artistic achievements. We could build a far more just and virtuous society. And there's no in-principle reason why civilisation couldn't reach other planets, of which there are some 100 billion in our galaxy.²⁵ If we let civilisation end, then none of this can ever happen.

We're unsure whether this great future will really happen, but that's all the more reason to keep civilisation going so we have a chance to find out. Failing to pass on the torch to the next generation might be the worst thing we could ever do.

So, a couple of percent risk that civilisation ends seems likely to be the biggest issue facing the world today. What's also striking is just how neglected these risks are.

Why these risks are some of the most neglected global issues

Here is how much money per year goes into some important causes:²⁶

Cause	Annual targeted spending from all sources (highly approximate)
Global R&D	\$1.5 trillion
Luxury goods	\$1.3 trillion
US social welfare	\$900 billion
Climate change	>\$300 billion
To the global poor	>\$250 billion
Nuclear security	\$1-10 billion
Extreme pandemic prevention	\$1 billion
AI safety research	\$10 million

As you can see, we spend a vast amount of resources on R&D to develop even more powerful technology. We also expend a lot in a (possibly misguided) attempt to improve our lives by buying luxury goods.

Far less is spent mitigating catastrophic risks from climate change. Welfare spending in the US alone dwarfs *global* spending on climate change.



But climate change still receives enormous amounts of money compared to some of these other risks we've covered. We roughly estimate that the prevention of extreme global pandemics receives under 300 times less, even though the size of the risk seems about the same.

Research to avoid accidents from AI systems is the most neglected of all, perhaps receiving 100-times fewer resources again, at around only \$10m per year.

You'd find a similar picture if you looked at the *number of people* working on these risks rather than money spent, but it's easier to get figures for money.

If we look at scientific attention instead, we see a similar picture of neglect (though, some of the individual risks receive significant attention, such as climate change):



Credit: Nick Bostrom

Our impression is that if you look at political attention, you'd find a similar picture to the funding figures. An overwhelming amount of political attention goes on concrete issues that help the present generation in the short-term, since that's what gets votes. Catastrophic risks are far more neglected. Then, among the catastrophic risks, climate change gets the most attention, while issues like pandemics and AI are the most neglected.

This neglect in resources, scientific study and political attention is exactly what you'd expect to happen from the underlying economics, and are why the area presents an opportunity for people who want to make the world a better place.

First, these risks aren't the responsibility of any single nation. Suppose the US invested heavily to prevent climate change. This benefits everyone in the world, but only about 5% of the world's population lives in the US, so US citizens would only receive 5% of the benefits of this spending. This means the US will dramatically underinvest in these efforts compared to how much they're worth to the world. And the same is true of every other country.

This could be solved if we could all *coordinate* — if every nation agreed to contribute its fair share to reducing climate change, then all nations would benefit by avoiding its worst effects.

Unfortunately, from the perspective of each individual nation, it's better if *every other* country reduces their emissions, while leaving their own economy unhampered. So, there's an incentive for each nation to defect from climate agreements, and this is why so little progress gets made (it's a prisoner's dilemma).



And in fact, this dramatically understates the problem. The greatest beneficiaries of efforts to reduce catastrophic risks are *future* generations. They have no way to stand up for their interests, whether economically or politically.

If future generations could vote in our elections, then they'd vote overwhelmingly in favour of safer policies. Likewise, if future generations could send money back in time, they'd be willing to pay us huge amounts of money to reduce these risks. (Technically, reducing these risks creates a trans-generational, global public good, which should make them among the most neglected ways to do good.)

Our current system does a poor job of protecting future generations. We know people who have spoken to top government officials in the UK, and many want to do something about these risks, but they say the pressures of the news and election cycle make it hard to focus on them. In most countries, there is no government agency that naturally has mitigation of these risks in its remit.

This is a depressing situation, but it's also an opportunity. For people who *do* want to make the world a better place, this lack of attention means there are lots high-impact ways to help.

What can be done about these risks?

We've covered the scale and neglectedness of these issues, but what about the third element of our framework, solvability?

It's less certain that we can make progress on these issues than more conventional areas like global health. It's much easier to measure our impact on health (at least in the short-run) and we have decades of evidence on what works. This means working to reduce catastrophic risks looks worse on solvability.

However, there is still much we can do, and given the huge scale and neglectedness of these risks, they still seem like the most urgent issues.

We'll sketch out some ways to reduce these risks, divided into three broad categories:

1. Targeted efforts to reduce specific risks

One approach is to address each risk directly. There are many concrete proposals for dealing with each, such as the following:

1. Many experts agree that better disease surveillance would reduce the risk of pandemics. This could involve improved technology or better collection and aggregation of existing data, to help us spot new pandemics faster. And the faster you can spot a new pandemic, the easier it is to manage.
2. There are many ways to reduce climate change, such as helping to develop better solar panels, or introducing a carbon tax.
3. With AI, we can do research into the "control problem" within computer science, to reduce the chance of unintended damage from powerful AI systems. A recent paper, [Concrete problems in AI safety](#), outlines some specific topics, but only about 20 people work full-time on similar research today.
4. In nuclear security, many experts think that the deterrence benefits of nuclear weapons could be maintained with far smaller stockpiles. But, lower stockpiles would also reduce the risks of accidents, as well as the chance that a nuclear war, if it occurred, would end civilisation.

We go into more depth on what you can do to tackle each risk within our problem profiles:

1. [AI safety](#)



2. [Pandemic prevention](#)

3. [Nuclear security](#)

4. [Climate change](#)

We don't focus on naturally caused risks in this section, because they're much less likely and we're already doing a lot to deal with some of them. Improved wealth and technology makes us more resilient to natural risks, and a huge amount of effort already goes into getting more of these.

2. Broad efforts to reduce risks

Rather than try to reduce each risk individually, we can try to make civilisation generally better at managing them. The “broad” efforts help to reduce all the threats at once, even those we haven't thought of yet.

For instance, there are key decision-makers, often in government, who will need to manage these risks as they arise. If we could improve the decision-making ability of these people and institutions, then it would help to make society in general more resilient, and solve many other problems.

Recent research has uncovered lots of ways to improve decision-making, but most of it hasn't yet been implemented. At the same time, few people are working on the issue. We go into more depth in our write-up of [improving institutional decision-making](#).

Another example is that we could try to make it easier for civilisation to rebound from a catastrophe. The Global Seed Vault is a frozen vault in the Arctic, which contains the seeds of many important crop varieties, reducing the chance we lose an important species. Melting water [recently entered the tunnel leading to the vault](#) due, ironically, to climate change, so could probably use more funding. There are lots of other projects like this we could do to preserve knowledge.

Similarly, we could [create better disaster shelters](#), which would reduce the chance of extinction from pandemics, nuclear winter and asteroids (though not AI), while also increasing the chance of a recovery after a disaster. Right now, these measures don't seem as effective as reducing the risks in the first place, but they still help. A more neglected, and perhaps much cheaper option is to create [alternative food sources](#), such as those that can be produced without light, and could be quickly scaled up in a prolonged winter.

Since broad efforts help even if we're not sure about the details of the risks, they're more attractive the more uncertain you are. As you get closer to the risks, you should gradually reallocate resources from broad to targeted efforts ([read more](#)).

We expect there are many more promising broad interventions, but it's an area where little research has been done. For instance, another approach could involve improving international coordination. Since these risks are caused by humanity, they can be prevented by humanity, but what stops us is the [difficulty of coordination](#). For instance, Russia doesn't want to disarm because it would put it at a disadvantage compared to the US, and vice versa, even though both countries would be better off if there were no possibility of nuclear war.

However, it might be possible to improve our ability to coordinate as a civilisation, such as by improving foreign relations or developing better international institutions. We're keen to see more research into these kinds of proposals.

Mainstream efforts to do good like improving education and international development can also help to make society more resilient and wise, and so also contribute to reducing catastrophic risks. For instance, a better educated population would probably elect more enlightened leaders (cough), and richer countries are, all else equal, better able to prevent pandemics — it's no accident that Ebola took hold in some of the poorest parts of West Africa.



But, we don't see education and health as the best areas to focus on for two reasons. First, these areas are far less neglected than the more unconventional approaches we've covered. In fact, improving education is perhaps the most popular cause for people who want to do good, and in the US alone, receives 800 billion dollars of government funding, and another trillion dollars of private funding. Second, these approaches have much more diffuse effects on reducing these risks — you'd have to improve education on a very large scale to have any noticeable effect. We prefer to focus on more targeted and neglected solutions.

3. Learning more and building capacity

We're highly uncertain about which risks are biggest, what is best to do about them, and whether our whole picture of global priorities might be totally wrong. This means that another key goal is to learn more about all of these issues.

We can learn more by simply trying to reduce these risks and seeing what progress can be made. However, we think the most neglected and important way to learn more right now is to do global priorities research.

This is a combination of economics and moral philosophy, which aims to answer high-level questions about the most important issues for humanity. There are only a handful of researchers working full-time on these issues.

Another way to handle uncertainty is to build up resources that can be deployed in the future when you have more information. One way of doing this is to earn and save money. You can also invest in your career capital, especially your transferable skills and influential connections, so that you can achieve more in the future.

However, we think that a potentially better approach than either of these is to build a *community* that's focused on reducing these risks, whatever they turn out to be. The reason this can be better is that it's possible to grow the capacity of a community faster than you can grow your individual wealth or career capital. For instance, if you spent a year doing targeted one-on-one outreach, it's not out of the question that you would find one other person with relevant expertise to join you. This would be an annual return to the cause of about 100%.

Right now, we are focused on building the effective altruism community, which includes many people who want to reduce these risks. Moreover, the recent rate of growth, and studies of specific efforts to grow the community, suggest that high rates of return are possible.

However, we expect that other community building efforts will also be valuable. It would be great to see a community of scientists trying to promote a culture of safety in academia. It would be great to see a community of policymakers who want to try to reduce these risks, and make government have more concern for future generations.

Given how few people actively work on reducing these risks, we expect that there's a lot that could be done to build a movement around them.

In total, how effective is it to reduce these risks?

Considering all the approaches to reducing these risks, and how few resources are devoted to some of them, it seems like substantial progress is possible.

In fact, even if we only consider the impact of these risks on the *present* generation (ignoring any benefits to future generations), they're plausibly the top priority.

Here are some very rough and simplified figures, just to illustrate how this could be possible. It seems plausible to us that \$100 billion spent on reducing existential threats could reduce it by over 1% over the next century. A one percentage point reduction in the risk would be expected



to save about 100 million lives among the present generation (1% of about 10 billion people alive today). This would mean the investment would save lives for only \$1,000 per person.²⁷

Greg Lewis has made a more detailed estimate, arriving at a mean of \$9,200 per life year saved in the present generation (or ~\$300,000 per life).²⁸ There are also more estimates in the thread. We think Greg is likely too conservative, because he assumes the risk of extinction is only 1% over the next century, when our estimate is that it's several times higher. We also think the next billion dollars spent on reducing existential risk could cause a larger reduction in the risk than Greg assumes (note that this is only true if the billion were spent on the most neglected issues like AI safety and biorisk). As a result we wouldn't be surprised if the cost per present lives saved for the next one billion dollars invested in reducing existential risk were under \$100.

GiveWell's top recommended charity, Against Malaria Foundation (AMF), is often presented as one of the best ways to help the present generation and saves lives for around \$7,500 (2017 figures).²⁹ So these estimates would put existential threat reduction as better or in the same ballpark of cost-effectiveness as AMF for saving lives in the present generation — a charity that was specifically selected for being outstanding on that dimension.

Likewise, we think that if 10,000 talented young people focused their careers on these risks, they could achieve something like a 1% reduction in the risks. That would mean that each person would save 1,000 lives in expectation over their careers in the present generation, which is probably better than what they could save by earning to give and donating to Against Malaria Foundation.³⁰

In one sense, these are unfair comparisons, because GiveWell's estimate is far more solid and well-researched, whereas our estimate is more of an informed guess. There may also be better ways to help the present generation than AMF (e.g. policy advocacy).

However, we've also *dramatically* understated the benefits of reducing existential threats. The main reason to safeguard civilisation is not to benefit the present generation, but to benefit future generations. We ignored them in this estimate.

If we also consider future generations, then the effectiveness of reducing existential threats is orders of magnitude higher, and it's hard to imagine a more urgent priority right now.

Now you can either read some responses to these arguments, or skip ahead to practical ways to contribute.

Who *shouldn't* prioritise safeguarding the future?

The arguments presented rest on some assumptions that not everyone will accept. Here we present some of the better responses to these arguments.

You need to focus more on your friends and family

We're only talking about what the priority should be if you are trying to help people in general, treating everyone's interests as equal (what philosophers sometimes call "impartial altruism").

Most people care about helping others to some degree: if you can help a stranger with little cost, that's a good thing to do. People also care about making their own lives go well, and looking after their friends and family, and we're the same.

How to balance these priorities is a difficult question. If you're in the fortunate position to be able to contribute to helping the world, then we think safeguarding the future



should be where to focus. We list concrete ways to get involved in the next section. Otherwise, you might need to focus on your personal life right now, contributing on the side, or in the future.

You think the risks are much lower than we've argued

We don't have robust estimates of many of the human-caused risks, so you could try to make your own estimates and conclude that they're much lower than we've made out. If they were sufficiently low, then reducing them would cease to be the top priority.

We don't find this plausible for the reasons covered. If you consider all the potential risks, it seems hard to be confident they're under 1% over the century, and even a 1% risk probably warrants much more action than we currently see.

You think there's almost nothing more we can do about the risks

We rate these risks as less "solvable" than issues like global health, so expect progress to be harder per dollar. That said, we think their scale and neglectedness more than makes up for this, and so they end up more effective in expectation. Many people think effective altruism is about only supporting "proven" interventions, but that's a myth. It's worth taking interventions that only have a small chance of paying off, if the upside is high enough. The leading funder in the community now advocates an approach of "hits-based giving".

However, if you were *much* more pessimistic about the chances of progress than us, then it might be better to work on more conventional issues, such as global health.

Personally, we might switch to a different issue if there were two orders of magnitude more resources invested in reducing these risks. But that's a long way off from today.

A related response is that we're already taking the best interventions to reduce these risks. This would mean that the risks don't warrant a change in practical priorities. For instance, we mentioned earlier that education probably helps to reduce the risks. If you thought education was the *best* response (perhaps because you're very uncertain which risks will be most urgent), then because we already invest a huge amount in education, you might think the situation is already handled. We don't find this plausible because, as listed, there are lots of untaken opportunities to reduce these risks that seem more targeted and neglected.

Another example like this is that economists sometimes claim that we should just focus on economic growth, since that will put us in the best possible position to handle the risks in the future. We don't find this plausible because some types of economic growth increase the risks (e.g. the discovery of new weapons), so it's unclear that economic growth is a top way to reduce the risks. Instead, we'd at least focus on differential technological development, or the other more targeted efforts listed above.

You think there's a better way of helping the future

Although reducing these risks is worth it for the present generation, much of their importance comes from their long-term effects — once civilisation ends, we give up the entire future.

You might think there are other actions the present generation could take that would have very long-term effects, and these could be similarly important to reducing the risk of extinction. In particular, we might be able to improve the *quality* of the future by preventing our civilization from getting locked into bad outcomes permanently.



This is going to get a bit sci-fi, but bear with us. One possibility that has been floated is that new technology, like extreme surveillance or psychological conditioning, could make it possible to create a totalitarian government that could never be ended. This would be the *1984* and *Brave New World* scenario respectively. If this government were bad, then civilisation might have a fate *worse* than extinction by causing us to suffer for millennia.

Others have raised the concern that the development of advanced AI systems could cause terrible harm if it is done irresponsibly, perhaps because there is a conflict between several groups raising to develop the technology. In particular, if at some point in the future, developing these systems involves the creation of sentient digital minds, their wellbeing could become incredibly important.

Risks of a future that contains an astronomical amount of suffering have been called “s-risks”.³¹ If there is something we can do today to prevent an s-risk from happening (for instance, through targeted research in technical AI safety and AI governance), it could be even more important.

Another area to look at is major technological transitions. We’ve mentioned the dangers of genetic engineering and artificial intelligence in this piece, but these technologies could also create a second industrial revolution and do a huge amount of good once deployed. There might be things we can do to increase the likelihood of a good transition, rather than decrease the risk of a bad transition. This has been called trying to increase “existential hope” rather than decrease “existential risk.”³²

We agree that there might be other ways that we can have very long-term effects, and these might be more pressing than reducing the risk of extinction. However, most of these proposals are not yet as well worked out, and we’re not sure about what to do about them.

The main practical upshot of considering these other ways to impact the future, is that we think it’s *even more* important to positively manage the transition to new transformative technologies, like AI. It also makes us keener to see more global priorities research looking into these issues.

Overall, we still think it makes sense to first focus on reducing existential threats, and then after that, we can turn our attention to other ways to help the future.

One way to help the future we *don’t* think is a contender is speeding it up. Some people who want to help the future focus on bringing about technological progress, like developing new vaccines, and it’s true that these create long-term benefits. However, we think what most matters from a long-term perspective is where we end up, rather than how fast we get there. Discovering a new vaccine probably means we get it earlier, rather than making it happen at all.

Moreover, since technology is also the cause of many of these risks, it’s not clear how much speeding it up helps in the short-term.

Speeding up progress is also far less neglected, since it benefits the present generation too. As we covered, over 1 trillion dollars is spent each year on R&D to develop new technology. So, speed-ups are both less important and less neglected.

To read more about other ways of helping future generations, see Chapter 3 of [*On the Overwhelming Importance of Shaping the Far Future*](#) by Dr Nick Beckstead.

You're confident the future will be short or bad

If you think it’s virtually guaranteed that civilisation won’t last a long time, then the value of reducing these risks is significantly reduced (though perhaps still worth taking to help the present generation and any small number of future generations).

We agree there’s a significant chance civilisation ends soon (which is why this issue is so important), but we also think there’s a large *enough* chance that it could last a very long time, which makes [the future worth fighting for](#).



Similarly, if you think it's likely the future will be more bad than good (or if we have much more obligation to reduce suffering than increase wellbeing), then the value of reducing these risks goes down. We don't think this is likely, however, because people *want* the future to be good, so we'll *try* to make it more good than bad. We also think that there has been significant moral progress over the last few centuries (due to the trends noted earlier), and we're optimistic this will continue. See more discussion in footnote 11.¹¹

What's more, even if you're not sure how good the future will be, or suspect it will be bad in ways we may be able to prevent in the future, you may want civilisation to survive and keep its options open. People in the future will have much more time to study whether it's desirable for civilisation to expand, stay the same size, or shrink. If you think there's a good chance we will be able to act on those moral concerns, that's a good reason to leave any final decisions to the wisdom of future generations. Overall, we're highly uncertain about these big-picture questions, but that generally makes us *more* concerned to avoid making any irreversible commitments.³³

Beyond that, you should likely put your attention into ways to decrease the chance that the future will be bad, such as avoiding s-risks.

You're confident we have much stronger moral obligations to help the present generation

If you think we have much stronger obligations to the present generation than future generations (such as person-affecting views of ethics), then the importance of reducing these risks would go down. Personally, we don't think these views are particularly compelling.

That said, we've argued that even if you ignore future generations, these risks seem worth addressing. The efforts suggested could still save the lives of the present generation relatively cheaply, and they could avoid lots of suffering from medium-sized disasters.

What's more, if you're uncertain about whether we have moral obligations to future generations, then you should again try to keep your options open, and that means safeguarding civilisation.

Nevertheless, if you think that we don't have large obligations to future generations *and* that the risks are also relatively unsolvable (or that there is no useful research to be done), then another way to help present generations could come out on top. This might mean working on global health or mental health. Alternatively, you might think there's another moral issue that's more important, such as factory farming.



Want to help reduce existential threats?

Our generation can either help cause the end of everything, or navigate humanity through its most dangerous period, and become one of the most important generations in history.

We could be the generation that makes it possible to reach an amazing, flourishing world, or that puts everything at risk.

As people who want to help the world, this is where we should focus our efforts.

If you want to focus your career on reducing existential threats and safeguarding the future of humanity, we want to help. We've written an article outlining your options and steps you can take to get started.

[HOW TO USE YOUR CAREER TO REDUCE EXISTENTIAL THREATS](#)

Once you've read that article, or if you've already thought about what you want to do, consider talking to us one-on-one. We can help you think through decisions and formulate your plan.

[APPLY FOR ONE-ON-ONE ADVISING](#)

Learn more

- Read about [51 policy and research ideas](#) for reducing existential risk.
- See the academic version of this argument in [this paper](#) by Professor Nick Bostrom.
- [Read the case for focusing on future generations.](#)
- Read an [alternative introduction to the idea that our impact on future generations is hugely morally important](#) — including a section on [objections](#).
- You can hear related ideas discussed in our podcasts with [Dr Toby Ord](#) and [Dr Nick Beckstead](#), as well as [other episodes](#) on specific risks.
- Listen to this interview with Thomas Moynihan on the [intellectual history of existential risk](#).
- We also recommend our podcasts with [Carl Shulman on the common-sense case for existential risk work and its practical implications](#) and with [Will MacAskill on what we owe the future](#).

80,000
HOURS

Read next:

This could be the most important century

Why the development of advanced AI systems that lead to explosive growth and scientific advancement may lead us more quickly than most people imagine to a deeply unfamiliar future — making the 21st century the most important in history.

[CONTINUE →](#)

Enter your email and we'll mail you a book (for free).

Join our newsletter and we'll send you a free copy of *The Precipice* — a book by philosopher Toby Ord about how to tackle the greatest threats facing humanity.

Email address

GET THE FREE BOOK

[EXISTENTIAL RISK](#) [FUTURE GENERATIONS & LONGTERMISM](#)

[MOST PRESSING PROBLEMS](#) [PROMISING INTERVENTIONS](#)

Notes and references

1. In the course of the last four months it has been made probable — through the work of Joliot in France as well as Fermi and Szilárd in America — that it may become possible to set up a nuclear chain reaction in a large mass of uranium, by which vast amounts of power and large quantities of new radium-like elements would be generated. Now it appears almost certain that this could be achieved in the immediate future.

This new phenomenon would also lead to the construction of bombs, and it is conceivable — though much less certain — that extremely powerful bombs of a new type may thus be constructed. A single bomb of this type, carried by boat and exploded in a port, might very well destroy the whole port together with some of the surrounding territory. However, such bombs might very well prove to be too heavy for transportation by air.



Einstein–Szilárd letter, Wikipedia,

[Archived link](#), retrieved 17-October-2017↵

2. Nick Bostrom [defines an existential risk](#) as an event that “could cause human extinction or permanently and drastically curtail humanity’s potential.” An existential risk is distinct from a [global catastrophic risk \(GCR\)](#) in its scope — a GCR is catastrophic at a global scale, but retains the possibility for recovery. An existential threat [seems to be used](#) as a linguistic modifier of a threat to make it appear more dire.↵
3. Greenberg surveyed users of Mechanical Turk, who tend to be 20-40 and more educated than average, so the survey doesn’t represent the views of all Americans. See more detail in this video:

[Social Science as Lens on Effective Charity: results from four new studies – Spencer Greenberg](#), 12:15.

The initial survey found a median estimate of the chance of extinction within 50 years of 1 in 10 million. Greenberg did three replication studies and these gave higher estimates of the chances. The highest found a median of 1 in 100 over 50 years. However, even in this case, 39% of respondents still guessed that the chances were under 1 in 10,000 (about the same as the chance of a 1km asteroid strike). In all cases, over 30% thought the chances were under 1 in 10 million. You can see a summary of all the surveys [here](#).

Note that when we asked people about the chances of extinction with no timeframe, the estimates were much higher. One survey gave a median of 75%. This makes sense — humanity will *eventually* go extinct. This helps to explain the discrepancy with some other surveys. For instance, “Climate Change in the American Mind” (May 2017, [archived link](#)), found that the median American thought the chance of extinction from climate change is around 1 in 3. This survey, however, didn’t ask about a specific timeframe. When Greenberg tried to replicate the result with the same question, he found a similar figure. But when Greenberg asked about the chance of extinction from climate change in the next 50 years, the median dropped to only 1%. Many other studies also don’t correctly sample low probability estimates — people won’t typically answer 0.00001% unless presented with the option explicitly.



However, as you can see, these types of surveys tend to give very unstable results. Answers seem to vary on exactly how the question is asked and on context. In part, this is because people are very bad at making estimates of tiny probabilities. This makes it hard to give a narrow estimate of what the population in general thinks, but none of what we've discovered refutes the idea that a significant number of people (say over 25%) think the chances of extinction in the short term are very, very low, and probably lower than the risk of an asteroid strike alone. Moreover, the instability of the estimates doesn't seem like reason for confidence that humanity is rationally handling these risks.↵