Starter guide

Your career is your biggest opportunity to make a difference. But how can you make the most of it? We've spent the last ten years searching for the answer to that question.

We've found that some paths open to you probably have far more impact than others, but they're often not what people are already focused on. This means we need to rethink social impact careers – and that by doing so, there's a chance you can find a career that's both higher-impact and personally satisfying.

In the series, you'll learn about the ideas that have most changed our view of what makes for a high-impact career. We start with how to define impact, and then go on to four key drivers of your impact, which you can use to compare your options in terms of the difference they make. Finally, we introduce the most important elements of career strategy.

There are 16 key articles, which you can read over a weekend. At the end, you'll have covered everything you need to know to start making a new career plan.

All this might sound like a lot of work, but if you can increase the impact of your career by just 1%, it would be worth spending up to 800 hours learning how to do that. And we think we can increase your impact by much more, and in much less time.

80,000 Hours is part of the Centre for Effective Altruism, a registered charity in England and Wales, Registered Charity Number 1149828
Centre for Effective Altruism, Oxford Uehiro Centre for Practical Ethics, Littlegate House, St Ebbes Street, Oxford OX1 1PT, UK

WWW.80000HOURS.ORG

Table of contents

1. Introduction

Summary: what makes for a high-impact career?

TL;DR: Get good at something that lets you effectively contribute to neglected global problems.

Prefer to listen rather than read? *Hear our founder and CEO, Ben Todd, discuss some of the most important ideas in the series in episode #71 of the 80,000 Hours Podcast.*

You can have more positive impact over the course of your career by aiming to:

1. Help solve a more pressing problem.
   Many global issues should get more attention, but as individuals we should look for the biggest gaps in existing efforts. To do that, you can compare issues in terms of scale, neglectedness, and tractability. It turns out that some issues receive *hundreds* of times less attention relative to how big and solvable they seem. This means which issues you choose to work on is likely the biggest driver of your impact. In particular, our generation may see the rise of transformative technologies, which could lead to existential risks and make now a crucial moment in history — but our current institutions are doing little to address these issues. We have a list of global issues we think are particularly pressing for more people to work on right now.

2. Find a more effective solution.
   Many social interventions don't have much effect when rigorously measured, but some are enormously effective. This means that by finding a more effective solution to your chosen problem, you can often make 10 or 100 times as much progress per year of work. To find these solutions, we advocate a 'hits-based approach': find rules of thumb that increase the chance of the solution being among the most effective in an area, even if it also has a good chance of not working. This often means working on research, policy change, or movement building.

3. Find a path with more leverage.
   Your 'leverage' is how many resources (e.g. money, attention, skill) you're able to mobilise toward the solution. To get more leverage on the most pressing problems, we often encourage people to work in government and policy; to pursue careers where they can mobilise others (e.g. media); to help people or organisations that have a lot of leverage; or to use their strengths to contribute

indirectly through donations or community building. Most people only reach their peak productivity between the ages of 40–60, so we also encourage people to invest in their skills, connections, reputation etc. to have more leverage in the future. See our list of especially promising career paths.

4. Find work that fits you better.
   The most productive people in a field often have far more output than the average. Plus, excelling in almost any field gives you more connections, resources and reputation, which give you leverage. So, once you have identified some promising options, choose between them based on your expected fit.

Career strategy: Your most impactful career is the one that's best on the product of these four factors over its course — and it's often possible to find a path that's 10 times better on one or more of these dimensions while being just as personally satisfying.

But how do you actually find the best possible path? Think like a scientist: make some best guesses at the most promising long-term paths, identify key uncertainties, then update your guesses every 1–3 years. Over time, especially focus on the following three (overlapping) stages:

1. Explore: learn about and test out promising longer-term paths, until you feel ready to bet on one for a few years. It's hard to predict where you'll have the best fit, but some paths are much higher-impact than others, so it's worth exploring to make sure you don't miss a great one. (This is typically the key focus for people ages 18–24.)
2. Invest: take a bet on a longer-term path by building the career capital that will most accelerate you in it. It's normally better to aim a little too high than too low — but make sure you have a backup plan, so you can try another path if it doesn't work out. (Ages 25–35.)
3. Deploy: use the career capital you've built to support the most effective solutions to the most pressing problems at the time. (Age 36 onwards.)

While doing this, seek community. Finding a great community gives you hundreds of connections at once, and two people working together can have more than twice the impact than one person alone. We helped to build the effective altruism community to help you (and ourselves!) find like-minded people.

Focusing your career on tackling the world's most pressing problems is not for everyone and is certainly not easy. If you're not able to change jobs right now, you can still have a lot of impact by enabling others, politically supporting, or donating to work on the problems you think are most pressing, while also investing in your career capital.

If you do change careers, look for work you enjoy and that meets your other needs. If a path feels like a struggle, it's probably not sustainable or inspiring to others, and so probably not ideal even for your impact.

Fortunately, we think the steps we recommend — building career capital, exploring, and contributing to meaningful problems — align with what's personally rewarding for a lot of people. And so while there may be some tradeoffs, we think these steps are a route to a career that's satisfying and fulfilling, as well as one that addresses some of the biggest issues of our time.

Example: Sophie Rose

Sophie was on a path to become a clinical doctor in Australia when she came across our research into how many lives a doctor saves in rich countries. She realised she might be able to have even more impact doing something else.

She heard our podcast with Beth Cameron about why a global pandemic is a realistic possibility (back in 2017!), though the issue was neglected compared to conventional medicine. So she decided to try to switch which problem she focused on.

She spoke to us one-on-one, and we helped her find funding for a master's in epidemiology to build career capital in the area.

When COVID-19 broke out, she was able to zero in on a neglected but potentially high-impact solution: human challenge trials, which can greatly speed up the development of vaccines.

She increased her leverage by co-founding 1DaySooner with a grant from others in our community. 1DaySooner is a nonprofit that signed up 30,000 volunteers for human challenge trials in order to speed up government vaccine approval — potentially enabling many academic labs to go ahead with studies. A challenge trial started in London in early 2021, setting a precedent that will hopefully make our response much faster if another pandemic breaks out.

Although Sophie would likely also have enjoyed being a doctor, this path has also been a good fit for her skills and because she thinks she's having a greater impact, she finds it meaningful. She now plans to double down on what she's learned, and work on preventing the (unfortunately realistic) possibility of a future pandemic even worse than COVID-19.

You can hear more about Sophie's story in this interview in Marie Claire.

What research is this based on?

Careers decisions are highly individual, so there are many questions we can't easily help with. We aim to focus on career questions that are more widely relevant. To answer the questions we tackle, we draw on:

- Expert interviews — you can listen to over 60 examples of these interviews on our podcast, and also see the results of some anonymous interviews, and our annual survey. Our first pass on many questions involves synthesising what several experts say on the question.
- Academic literature — we aim to draw on academic literature where it's available, such as the literature on existential risks, the distribution of productivity in different fields, and how to make good decisions. We're also affiliated with some academic partners.
- Advising our readers — we've given one-on-one advice to over 1,000 people since 2011, many of whom we're still in touch with. This gives us a sense of what mistakes are common, as well as some indication of how decisions play out over time.

It's not usually possible to confidently answer the kinds of questions we tackle. However, we do our best to synthesise the sources of evidence we draw on, using our research principles. We also aim to highlight the key aspects of our reasoning so that readers can make their own assessments.

Advice on how to read our advice

The topics we tackle are complex, and in the past we've noticed people interpreting our advice in ways we didn't intend. Here are some points to bear in mind before diving into our advice.

- We want our writing to inform people's views, but only in proportion to the likelihood that we are correct. Given that, it's important to keep in mind that **we've been wrong before and we'll be wrong again**. We've spent a lot of time thinking about these issues, but we still have a lot to learn. Our positions often change every couple of years, and due to the nature of the questions we take on we're rarely more than about 70% confident in our answers. You should try to strike a balance between what we think and your previous position, depending on the strength of the arguments and how much you already knew about the topic.
- **It's extremely difficult to give universally applicable career advice**. The most important issue here is that which option is best for you depends a huge amount on your skills and circumstances, and the specific details of the opportunity. So, while we might highlight path A more than path B, the best opportunities in path B will often be better than the typical opportunities in path A. Moreover, your personal circumstances could easily mean the best option for you is in path B. So, treat the specific options we mention as an aid for compiling your personal list of career ideas. Also keep in mind that many issues in career choice are a matter of balancing opposing considerations — for instance, if we say people put too much emphasis on X, there will usually be *some* readers who put too little emphasis on X, and need to hear the opposite advice.
- **Our advice is aimed at a particular audience**: namely, people with college degrees who want to make having a positive impact (from an impartial perspective) the main focus of their careers, especially in the problem areas we most recommend; who live in rich, (for the most part) English-speaking countries; and who want to take an analytical approach to their career. At any given moment many people need to focus on taking care of their own lives, and we don't think anyone should feel guilty if that's the case. Certain parts of our advice, such as our list of priority paths, are especially aimed at people who are unusually high achieving. In general, the more similar you are to our core audience, the more useful the advice will be, although much of what we write is useful to anyone who wants to make a difference.
- **Treat increasing your impact as just one long-term goal**. Working on the world's most pressing problems is among the most worthwhile challenges we can imagine, though it can also be overwhelming. Bear in mind, 80,000 Hours is about how to maximise your impact, and this can make it sound like we don't

care about other goals. However, the team sees increasing our impact as just one important goal among several in our lives, which means we often do things that aren't optimal from the perspective of doing good. Indeed, even if your *only* goal was to have an impact, to do that it's vital to do something you can stick with for years — and this means taking care of your personal priorities as well.

- Aim for steady progress rather than perfection. It can take a long time to work out how to factor the ideas we cover into your own plans and find the right opportunity. Along the way, because there's always more that could be done, it can be easy to become overly perfectionist, get caught up with comparisons, and never be satisfied. When using our advice, the aim is not to find the (unknowable and unattainable) perfect option, or have more impact than other people. Rather, focus on making steady progress towards the best career that's practical for you given your constraints.
- Older articles on the site are less likely to reflect our current views, so check their publication date. We also aim to keep this key ideas page up to date as the canonical source of advice, and to flag older articles when our views have changed, though we have hundreds of pages of content, so we don't catch everything.

  Read more.

Now, you can read the full key ideas series below to learn more about each step, or go straight to making a career plan.

This is your most important decision

Click to read online

You have about 80,000 hours in your career: 40 hours per week, 50 weeks per year, for 40 years.

This makes your choice of career the most important ethical decision of your life.

From adulthood, you'll spend almost half your waking hours on your career — more than the time you'll spend eating, on hobbies, and watching Netflix put together. So unless you happen to be the heir to a large estate, that time is the biggest resource you have to help others.

This means if you can increase the positive impact of those hours just a little, it will probably have a bigger impact than changes to any other parts of your life.

And we're going to show that — if you have the privilege to have options — you can probably increase the impact of your career not just by a little, but by a huge amount.

Some of the paths open to you probably do vastly more for the world than others, but they're probably not the ones you're currently focusing on.

Our generation faces issues of historical importance, and you can help tackle them. And you don't have to become a doctor, teacher or a charity worker — there are many other routes to contribute that can be even higher impact.

By considering a broader range of routes, you may be able to find a path that has far more impact, while being just as — or even more — fulfilling as the path you're currently on.

So how did we end up thinking this?

When our founders, Ben and Will, were about to graduate in 2011, they wanted to find work they enjoyed, that paid the bills, and that made a contribution to the world. But they felt really unsure what to do.

Given the huge amount of time at stake, it seemed worth doing some real research to find the best path. But it didn't feel like the existing careers advice even *tried* to compare paths in terms of impact, nor was it based on much evidence. So they started to research the question themselves.

After 10 years of research, we now think the impact you can have in different career paths — however exactly you define 'impact' — is driven by four main factors:

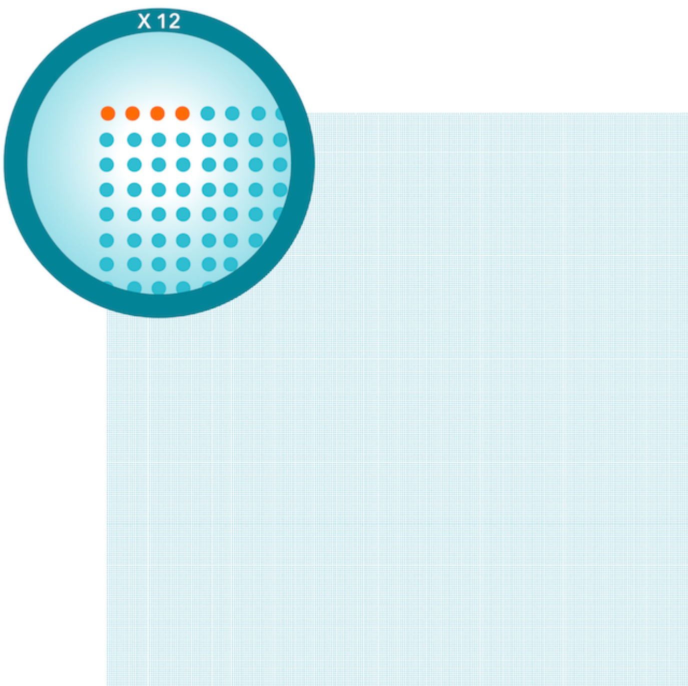
1. How **pressing the problems** you focus on are

2. How **effective the solutions** you pursue are
3. The amount of **leverage** you can apply to those solutions
4. Your **personal fit** for the path

In this article, we'll show how you can use these factors to compare your options in terms of impact, and find new ways to contribute. (If you just want to see the bottom lines, here's a two page summary.)

We'll also explain how the rest of our (free) research and support can help you apply these ideas and give you the best possible chance of finding a rewarding career that fulfils your potential for impact.

Understanding and applying these ideas takes time. But if you'd spend six minutes discussing where to go for a two-hour dinner, you should be willing to spend up to 4,000 hours researching your career. And while we sometimes do go on a bit, we promise it'll be quicker than that.



Each dot illustrates one of the 80,000 hours in your career.
You can read our key ideas in under four of them.

So let's start with the first and most important factor.

Which problems should you work on?

We think the most crucial question you face in choosing an impactful career is which problem or issue to focus on — whether that's health, education, climate change, engineered pandemics, or something else.

Most people seem to think that comparing the importance of these kinds of issues is near impossible, and you should just work on whatever issue you're passionate about.

But we think it's clear that some problems are much bigger than others.

Climate change is widely considered one of the world's biggest problems, and we think it's even bigger than often supposed. While the most likely scenario is several degrees of warming, the uncertainty in climate models means it's hard to rule out a 5% chance of over 10°C warming.

What's more, the CO2 we emit today will stay in the atmosphere for tens of thousands of years, impacting our children's grandchildren and beyond. We think future generations matter, which makes the issue even bigger in scale.

But we also think there may be issues that are even larger again.

The philosopher Toby Ord has argued that in 1945 humanity entered a new age, which he calls *the Precipice*. On July 16, 1945 humanity detonated the first atomic bomb, which would eventually make it possible — for the first time in history — for a small group of people to destroy most of the world's cities within hours.

The annual risk of an all-out nuclear exchange is small, but it's not zero: more countries are armed than ever, and there is always the chance of an accident or malfunction. The average of several expert surveys is an annual probability of 0.4%.

Compounded over our lives and the lives of our children, this adds up to a substantial chance of a catastrophe potentially more devastating than climate change.

Besides killing most people living in urban areas, the resulting fires could lift enough ash into the air to obscure the sun and reduce global temperatures for years, leading to widespread famine.

Within months, this would not only kill most people alive today, but it could also lead to a collapse of civilisation itself. We think a permanent collapse is very unlikely, but when we consider the scale of the consequences — the loss of all future generations — that risk may be the worst thing about a nuclear conflict.

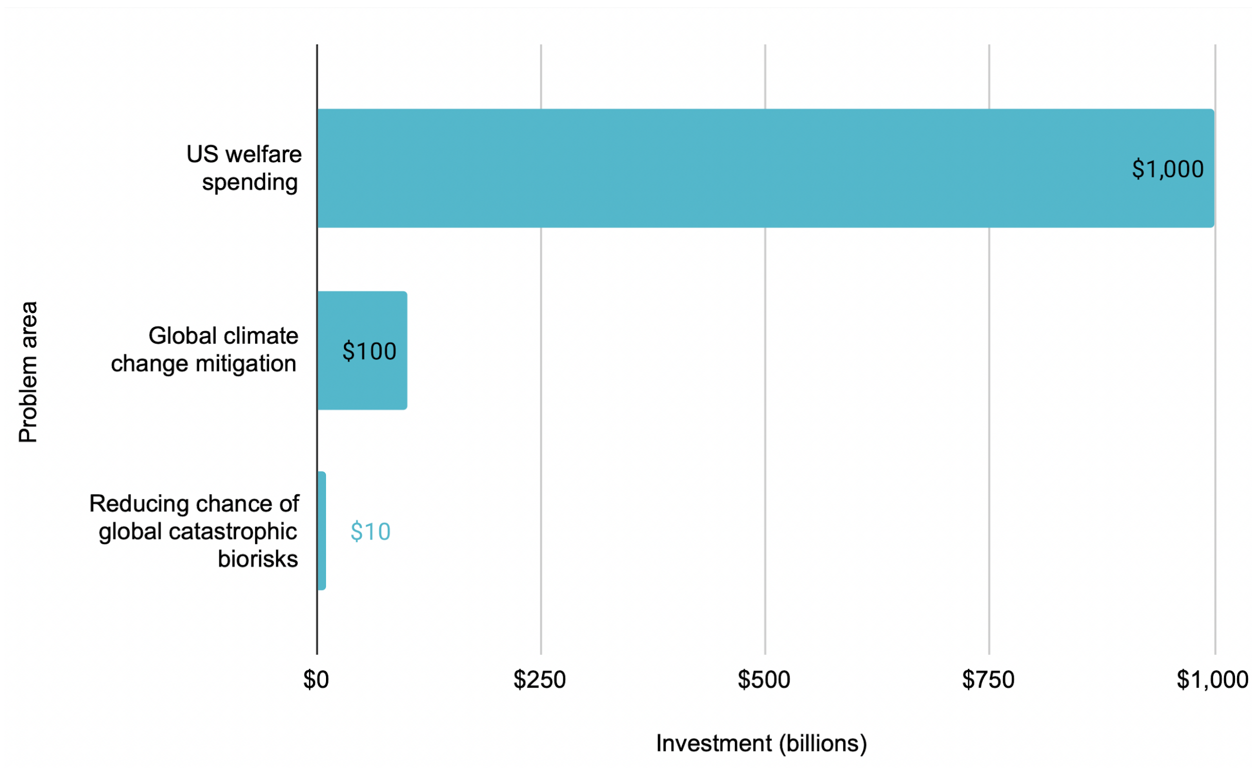
But the possibility of extreme climate change and nuclear winter are just two examples of a broader trend.

Technology has given this generation unprecedented power to shape history. The consequences of the decisions we make today about nuclear weapons, genetic engineering, artificial intelligence, space settlement, and other emerging technologies could ripple forward for thousands of years, with either enormously positive or enormously negative consequences.

So: Some problems are *much bigger* than others.

At the same time, some problems are much more **neglected** by society.

For instance, in 2016 we argued that a global pandemic posed a significant global risk, but with around $10 billion spent globally on preventing the worst pandemics,[1] it was more neglected than climate change or international development (which receive hundreds of billions) — which are in turn far more neglected than education and health in rich countries (which receive trillions).



This is important because the 1,000th person working on an issue will most likely drive much less progress than the 10th person, an effect that economists call diminishing returns.

[1] Greg Lewis estimates that a quality-adjusted ~$1 billion is spent annually on global catastrophic biorisk (GCBR) reduction. Most of this comes from work that is not explicitly targeted at GCBRs, but is rather disproportionally useful for reducing them. The US budget for health security in general is ~$14 billion. Worldwide, the budget is probably something like double or triple that — so spending that's particularly helpful for GCBR reduction is probably just a few percent of the total; the spending for explicit GCBR reduction would be much less. See the relevant section of our GCBR profile, including footnote 21.

As individuals, our role should be to find the biggest gaps in the priorities served by the current economic and political system, and do our best to fill them.

Inspired by this argument, one of our readers, Cassidy, realised that the knowledge she'd learned as a doctor might be put to even better use in pandemic prevention.

She applied to a masters in public health programme, and from there was able to get into a biosecurity PhD at Oxford. Since the field of pandemic control was relatively small before COVID-19, she was able to advance quickly. When COVID-19 broke out, she was ready to advise the UK government on policy ideas to control COVID-19, as well as how to prevent the next (potentially much worse) outbreak.

Cassidy realised she could help more people by working on pandemic policy than by seeing one patient at a time as a doctor.

The importance of working on neglected issues means that following your current passions could easily point you in the wrong direction. You're most likely to stumble across the same issues everyone is already talking about, which will usually be the *least* neglected. The best options are probably unconventional.

Which issues might be even more neglected than pandemics?

A survey of researchers found that they believe there's about a 50% chance that AI systems will exceed human capacities in most jobs around 2060. This would be one of the most important events in history.

These same researchers also estimated that if this happened, while the outcome might be 'extremely good,' there is also a 5% chance the outcome could be 'extremely bad' (e.g. human extinction). And as the people developing the technology, they're probably on the optimistic side.

One reason some are concerned is that it's unclear that as AI systems become more powerful, we can guarantee they'll continue to stay aligned with human values. This could suggest that research into 'AI alignment' is a crucial challenge from a long-term perspective.

Since we first wrote about it in 2012, AI alignment has grown into a flourishing field of research in computer science, but it still receives under $100 million in funding per year — about 100 times less than preventing pandemics.

Long-term AI *policy* — addressing the question of how society and government should handle advancing AI — is yet more neglected. (Read more about our case for both.)

You might be surprised at how your current skills can be applied to unusual problems like AI alignment (and later we'll explain how to contribute no matter your current job).

Brian Tse was working at an investment bank in Hong Kong and didn't have a background in AI. But he started to learn more about it in his spare time, and using his bilingual background, started to help translate materials to connect Western and Chinese researchers.

He now runs an independent consulting firm advising organisations in China, the United States, and Europe on the safety and governance of AI and other potentially transformative technologies — a path he also finds much more fulfilling and exciting than banking.

While working in the corporate world, Brian learned about AI and other transformative technologies. He now consults with many key Chinese and Western organisations that work on AI, with the aim of helping them coordinate better.

Comparing global problems is, of course, profoundly difficult. So we've also supported the development of the field of global priorities research, which tries to break down and answer the questions involved — for example, how we should allocate resources between reducing the risk of a disaster, on the one hand, and preventing a more certain harm on the other. There's now an institute dedicated to this topic at Oxford. But despite the vital importance of this research, there are still only tens of researchers directly focused on it.

Ben Garfinkel was considering a physics PhD, but thought that since so many extremely smart people already do physics research, it would be hard to make a big contribution. He learned about global priorities research through a talk we gave, and after testing it out, decided to switch fields. He then developed important criticisms of arguments for prioritising AI.

This kind of criticism is exactly what we want to see more of. There's a good chance our views of which problems are most pressing are mistaken in some ways or incomplete, and we want people to make the case for alternatives, and find issues we haven't even thought of yet.

You don't have to decide on which problem to focus on right away. As we cover in our advice on career strategy and planning, early in your career it often makes sense to focus more on exploring and building skills. However, since your choice of problem is such an important driver of your impact, it's vital to think about as you go on, and especially before you get locked into a path.

On our site, you can find a list of ideas for global problems to work on, including many issues we didn't have a chance to mention here. Some involve neglected ways to improve our economic and political system and build a better society in general, as well as ways to directly tackle risks to the future. We also have guidance on how to compare them, and profiles about how to contribute to solving each one.

What are the best solutions to those problems?

Whatever problem you decide to focus on, you need to choose an effective way to address it.

Scared Straight was a government programme that received billions of dollars of funding, and was made into an award-winning documentary. The idea was to take kids who committed crimes, show them life in jail, and scare them into embracing the straight and narrow.

The only problem? Two meta-analyses showed that the programme made kids *more* likely to commit crimes.[2] One study even estimated that every $1 spent on this programme caused $200 of *harm to society*.

Causing this much harm is rare, but when social programmes are rigorously tested, a large fraction of them don't work.[3] So it would be easy to end up working on a solution that has very little impact.

---

[2] van der Put, Claudia E., et al. "Effects of awareness programs on juvenile delinquency: a three-level meta-analysis." *International journal of offender therapy and comparative criminology* 65.1 (2021): 68-91. Archived link. Petrosino, Anthony, et al. "Scared Straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review." *Campbell Systematic Reviews* 9.1 (2013): 1-55. Archived link

[3] The percentage that work or don't work depends a lot on how you define it, but it's likely that a majority don't have statistically significant effects.

Meanwhile, research finds that among solutions that *are* effective, the best interventions are *far* more cost effective than average, often achieving 10 or 100 times as much for a given unit of resources.

For example, in recent years there's been a wave of advocacy to stop the use of plastic bags.

Focusing on advocacy is a route to having much more impact than just changing your personal consumption, but it's unclear this advocacy is well directed. Convincing someone to entirely give up plastic bags *for the rest of their life* (about 10,000 bags) would avoid ~0.1 tonnes of CO_2 emissions. In contrast, convincing someone to take just *one fewer* transatlantic flight would reduce CO_2 emissions by more than 10 times as much.[4]

And rather than trying to change personal consumption in the first place, we'd argue you could do even more to reduce emissions by advocating for greater funding of neglected green technology.

In general, it's vital to look for the *very best* solution in an area, rather than those that are just 'good'. This contrasts with the normal attitude that emphasises 'making *a* difference', rather than making *the most* difference possible.

In practice, this involves using rules of thumb to identify solutions that might make an especially big contribution if successful — even if there's a good chance of failure — as well as solutions that are unfairly neglected. We call trying to identify solutions that might be positive outliers the 'hits-based' approach.

Find out what experts believe are the best interventions in their fields by checking out our problem profiles and podcast interviews.

In which career can you apply the most leverage to those solutions?

Once you've picked a problem to focus on and identified some promising solutions, you need to choose a career that will let you make them happen. And some careers will let you contribute far more resources to great solutions than others. We call this your *leverage*.

People who want to have an impact often focus on jobs where they help people directly: teaching and healthcare are two of the most common paths for college graduates. But you might well be able to find options with more leverage.

---

[4] According to the 2020 Founders Pledge Climate & Lifestyle Report, just one roundtrip transatlantic flight contributes 1.6 tonnes of CO_2. Figure 2 of the same report shows the comparatively negligible effect of reusing plastic bags.

We worked with a doctor, Greg Lewis, to estimate the number of lives saved by a typical clinical doctor. While the impact is significant, it's less than commonly thought.

One reason is that the problem of health in rich countries already receives a lot of attention compared to issues like pandemics and nuclear security. But another reason is that a doctor can only treat a limited number of people each year. This motivated both Greg and Cassidy (from earlier) to switch from clinical medicine to public health and policy.

Another reader, Suzy Deuster, wanted to become a public defender to ensure disadvantaged people have good legal defence, but she realised that while that path might improve criminal justice for perhaps hundreds of people over her career, by changing policy she might improve justice for thousands or even millions. She was able to use her legal background to start a career in policy, and now works in the Executive Office of the President on criminal justice reform, and from there can explore other areas of policy in the future.

There are also other routes to leverage besides working in policy.

For instance, you can likely contribute more by building a community working on your chosen problem than you could achieve directly. If you can get just one other person to join you, then you've potentially doubled your impact. This means you can potentially have a very big impact no matter your current job.

The importance of spreading neglected ideas could also suggest pursuing careers in the media, or building any form of following. Isabelle Boemeke started out as a fashion model, but after speaking to experts who said nuclear energy was needed to tackle climate change, decided to use her platform to promote it as a neglected solution to climate change.

You can have leverage by helping *someone else* who has leverage, such as through working in operations or as a personal assistant. You can also help to build organisations which have leverage — whether nonprofit or for-profit. We list high-impact organisations hiring people with all kinds of skills on our job board.

Another form of leverage comes from money: Donations can be targeted at the most effective organisations in the world that are most in need of funding. Even if you have a very specific skill set — or don't want to change your career — by earning and donating money, you can 'convert' your labour into skilled labour working on the most pressing issues.

We call this 'earning to give' — finding a career that uses your strengths and allows you to donate more, even if its direct impact is only neutral.

An extreme example is Sam Bankman-Fried. He learned about the arguments for earning to give when he attended a talk by one of our founders while studying physics as an undergraduate at MIT.

Through others he met in our community, Sam found a job that used his mathematical skills in quantitative trading at Jane Street Capital, and that was a great fit. From there, he went on to help found cryptocurrency derivatives exchange FTX. Now, *Forbes* estimates his net worth is $16 billion, making him likely the world's wealthiest person under 30. Sam has already donated millions to causes like animal welfare and the Biden campaign, and intends to donate most of his future wealth.

Sam took earning to give to the extreme. Going from a physics undergrad to a pioneer in decentralised finance in less than a decade, Sam is now worth billions — which he plans to give away.

Over 500 of our readers are also pursuing earning to give on a more modest scale. For example, John Yan decided that a good way for him to contribute is to stay in his current job (software engineering) and donate 10–30% of his income to effective charities. Collectively the contributions of these readers will add up to tens of millions of dollars of donations, which can do a huge amount of good.

Finally, remember that leverage requires time. While many young people want to have a big impact right away, research finds that most people reach their peak output at age 40–60.[5]

This means that if you don't feel like you have much leverage now, another option is to build what we call your 'career capital': skills, connections, and contributors to your reputation that will let you have more influence in the future.

---

[5] Fatih Guvenen, Fatih Karahan, Serdar Ozkan, Jae Song. *What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?* Staff Report No. 710 February 2015. Archived link
Simonton, Dean K. "Age and outstanding achievement: What do we know after a century of research?" *Psychological Bulletin* 104.2 (1988): 251. Archived link
Shane Snow. "These are the ages when we do our best work", *Fast Company*, 2016, Archived link
Jones, Benjamin, E. J. Reedy, and Bruce A. Weinberg. *Age and scientific genius.* No. w19866. National Bureau of Economic Research, 2014. Archived link

In our guides, we cover how to compare careers in terms of leverage and which options are best for career capital, as well as a list of ideas for high-leverage paths and current open job opportunities. If you don't want to change your job right now, here's how to invest in yourself, and how to have a big impact in your current position.

Personal fit: what will you be good at — and what do you actually want to do?

"Find work you're good at" is a truism, but we think many people still don't take it seriously enough.

Data shows that in many fields, success is distributed very unevenly.

Data on the dispersion of staff productivity			
Field & outcome	**Source**	**Share of output from the top...**	
		20%	**1%**
Output in "low" complexity jobs among applicants e.g. mail carrier	Hunter, Schmidt, & Judiesch 1990	51%	3%
Output in "medium" complexity jobs among applications e.g. cook	Hunter, Schmidt, & Judiesch 1990	58%	4%
Papers coauthored by mathematicians with at least 133 publications	Clauset et al. 2009	33%	4%
Papers written by scientist (whole career)	Sinatra et al. 2016	39%	4%
Weeks in Billboard Hot-100 (1970-2018) by musician, among artists with at least 282 weeks in these charts	Tauberg 2018	35%	5%
Box Office Gross by US top-200 movie director	Tauberg 2018	40%	7%
Citations to scientists (whole career)	Sinatra et al. 2016	51%	7%
Income (worldwide, 2005)	Anand & Segal 2014		21%
Weeks on NYT Fiction Bestseller list by author with at least 6 weeks on that list	Tauberg 2018	76%	46%
Startup founder equity by company, among Y Combinator companies	80,000 Hours 2014		>80%

A top writer, for instance, could easily have 100 times the audience of the median. So even a small predictable difference in your suitability for two paths could translate into a big difference in outcomes.

Another reason is that being successful in almost any field gives you more connections, credibility, and money to direct towards neglected problems — increasing your leverage.

For both reasons, you need to balance your fit with a path with the other factors covered above, as well as how satisfying you'd find it. We'd rarely recommend doing something you dislike because it seems higher impact.

And if you might be exceptional at something, it could be worth pursuing, even if you're not sure how it'll be useful to pressing problems. As covered in the previous section, even if you work in finance or fashion, you can use your career to have a big impact.

So how can you predict what you'll be best at?

Despite the importance of this question, there doesn't seem to be much good evidence-based advice out there.

To help fill this gap, we created guides on how to predict your fit, what makes for a satisfying job, and how to plan your career over time.

Predictions can only go so far though, so while we'd encourage you to have hypotheses about which long-term paths will be best for you, we'd also encourage an iterative approach. Especially early in your career, you'll learn a huge amount about your fit and the other factors, so it's often best to find a good step for the next 1-5 years, then reassess your longer-term goals, then take another step, and repeat.

How much do careers differ in impact?

We've shown that you can have more impact by:

1. Finding a bigger and/or more neglected problem
2. Finding a more effective solution
3. Finding a path with more leverage
4. Finding work that fits you better

We've also shown that there are big differences for each.

But now also note that the differences *multiply together*, rather than merely add up.

For instance, if you can find a problem where additional resources are twice as effective, and — through greater leverage — direct twice as many resources to those problems, then you'll have four times as much impact.

If you can find a path that's twice as good on each dimension, it would be 16 times higher impact in total.

And we've shown that despite the huge uncertainties involved, it's plausible that some paths are at least 10 times higher impact than others on each dimension. This means that when multiplied together, the differences across all factors could be 100 fold or even 1,000 fold.

It's easy to gloss over these differences in scope, but let's appreciate for a moment just how much they matter. It could mean saving 100 times more lives, reducing carbon emissions 100 times as much, or making 100 times more progress reducing the biggest risks facing humanity.

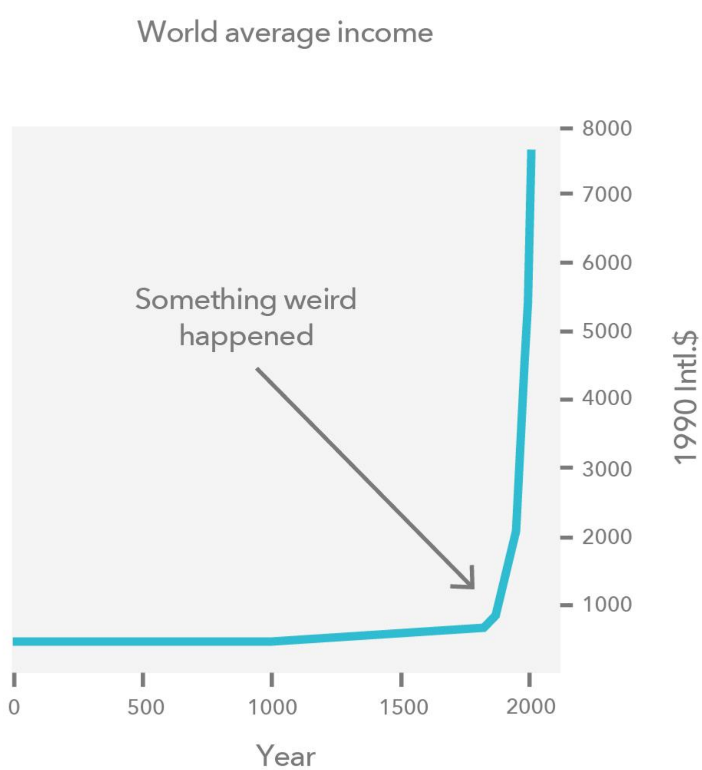
These differences are not the only ethically relevant factors, and everyone has priorities in life besides moral ones, but they do really matter.

Another implication is that if it's possible to find an option that's 100 times higher impact than your current best guess, then ten years in that path would achieve what could have otherwise taken people like you *1,000 years*. You could then spend the next 30 years on a beach doing whatever you like, and still have done far more to help others.

The potential spread also illustrates why career choice is such an important ethical decision. We associate 'ethical living' with things like recycling or reducing your CO2 emissions, but even with great effort, the best you could do is cut your personal footprint to near zero. But by thinking carefully about your career, you can affect the consumption of hundreds or even millions of people, and also contribute to even bigger and more neglected priorities.

How is it possible that such big differences in impact exist?

One reason is the massive economic and technological bounty of the industrial revolution, which means that today, many ordinary citizens of rich countries have what would have been kinglike wealth and power in previous centuries.



Advancing technology may make the next century one of the most important in history.

Our generation can wreck the climate for thousands of years, or we can build a sustainable economy. We can continue to expand factory farming, or we can eradicate it. We can allow new technologies like nuclear weapons to end civilisation, or we can usher in a future better than we can easily imagine — and be good stewards for all future generations.

Our aim is to help people like you understand this new power. If you have the good fortune to have options about how you spend your career, you can help change the course of history on these vital issues.

This is not an easy path, but it is a worthwhile one.

We've often felt racked with uncertainty about what to do, and overwhelmed at the scale of the issues. But we've also found a great deal of meaning and satisfaction in our efforts, especially as more and more people have united around them.

We still have a lot to learn, but we hope that by sharing what we've learned so far we can help you avoid the mistakes we've made, and speed you along your path to an impactful career.

How we can help you

We founded 80,000 Hours in 2011 to freely provide the information and support we wish we'd had when we graduated: transparently explained, based on the best research available, and willing to ask the big questions.

By doing this, we hope to get the next generation of leaders tackling the world's biggest and most neglected problems.

To date, millions of people have read our advice, and thousands of people have told us they've changed careers based on it.

Here's how you can get started right now, and improve your most important decision.

Understand how to increase your impact

Our key ideas guide aims to summarise all our research on the factors discussed in this essay (and more), in order to help you answer the key questions that determine your impact:

1. What does it mean to make a difference, and how can we know what helps?
2. Which global problems are most pressing?
3. Which solutions are most effective?

4. Which careers give you the most leverage?
5. Which jobs will you be good at?
6. How can you best invest in your skills?
7. What does a satisfying job look like?
8. What strategy should you take to pursue a great career?

You can also explore these ideas by subscribing to our podcast, which features in-depth conversations with experts on the world's most pressing problems and how you can help solve them.

Get ideas for high-impact career paths tackling the world's most pressing problems

We've reviewed different global problems, career paths, and jobs to help you find new ways to contribute. See:

1. A list of neglected global problems we think are especially pressing
2. Some broad types of high-leverage careers, and a list of specific career paths that we think best contribute to the problems we highlight
3. Ideas on how to apply your existing skills to have an impact
4. How to make a big difference without changing jobs

Due to our limited capacity, our specific recommendations are aimed at college students and graduates aged 18–30 who want to make impact their main focus — though the principles we cover apply to everyone.

Find jobs

We have a list of current job opportunities that can let you have a big impact, or build career capital for high-impact roles.

Join our newsletter, and we'll send you a curated list of the top opportunities about twice a month.

Make your career plan

When you have some ideas, check out our resources on planning and decision making, which aim to help you feel confident in your next decision.

Our 8-week career planning course helps you tie together everything you need to consider to make your plan and start taking action.

Get one-on-one advice

You can apply to speak to our team one-on-one for free. They can help check your plan, and may be able to make introductions to jobs and mentors who can help you get started.

Get help from our community

Two people working together can often achieve more than twice as much, so to help you find collaborators, we've also helped to build the effective altruism movement.

Effective altruism is the study of how best to help others. It tackles similar questions to those we've covered — which problems, solutions, and methods are most effective — and expands the focus to encompass all ways of doing good, whether choosing a career, donating, or engaging politically, while aiming to bring evidence and careful reasoning to bear on the questions.

Besides being a field of research, effective altruism is also a movement of people trying to put its findings into practice. There are now tens of billions of dollars committed to the approach, and thousands of people working together to use their careers to help others.

You can get involved online, through conferences in five continents, and in hundreds of local groups full of people keen to collaborate and help you have a greater impact. Through the community, we've found some of the most impressive and dedicated people we've ever met.

No time to read right now?

Join our newsletter and we'll send you a list of our most important articles, job opportunities, and monthly updates on new research.

You'll be joining our community of over 150,000 people, and can unsubscribe in one click.

2. Foundations: what does it mean to make a difference?

What is social impact? A definition

Click to read online

Lots of people say they want to "make a difference," "do good," "have a social impact," or "make the world a better place" — but they rarely say what they mean by those terms.

By clarifying your definition, you can better target your efforts, and make a difference more effectively.

But how *should* you define social impact?

Thousands of years of philosophy have gone into that question. We're going to try to sum up that thinking; introduce a practical, rough-and-ready definition of social impact; and explain why we think it's a good definition to focus on.

This is a bit ambitious for one article, so to the philosophers in the audience, please forgive the enormous simplifications. We hope the usefulness of the definition will make up for it.

A simple definition of social impact

If you just want a quick answer, here's the simple version of our definition (a more philosophically precise one — and an argument for it — follows below):

> *Your social impact is given by the number of people[6] whose lives you improve and how much you improve them, over the long term.*

This shows that you can increase your impact in two ways: by helping more people over time, or by helping the same number of people to a greater extent (pictured below).

---

[6] We often say "helping people" here for simplicity and brevity, but we don't mean just humans — we mean anyone with experience that matters morally — e.g. nonhuman animals that can suffer or feel happiness, even conscious machines if they ever exist.

We say "over the long term" because you can help more people either by helping a greater number now, or taking actions with better long-term effects.

This definition is enough to help you figure out what to aim at in many situations — e.g. by roughly comparing the number of people affected by different issues. But sometimes you need a more precise definition.

A more rigorous definition of social impact

Here's our working definition of "social impact":

> *"Social impact" or "making a difference" is (tentatively) about promoting total expected wellbeing — considered impartially, over the long term — without sacrificing anything that might be of comparable moral importance.*

In the rest of this article, we'll expand on:

1. Why we think social impact is primarily about 'promoting' what's of value — i.e. making the world better — rather than other kinds of moral considerations.
2. Why we think making the world a better place is in large part about promoting total expected wellbeing.
3. What we mean by "without sacrificing anything that might be of comparable moral importance."

4. How we can assess what makes a difference in the face of uncertainty.

We believe that taking this definition seriously has some potentially radical implications about where people who want to do good should focus, which we also explore.

If you're interested to understand the philosophy behind these ideas and their implications, read on. Otherwise, you can skip to the next article in our key ideas series.

Two final notes before we start. First, our definition is tentative — in that there's a good chance we're wrong and it might change. Second, its purpose is practical — it aims to cover the *most important* aspects of doing good to help people make better real-life decisions, rather than capture everything that's morally relevant.

In a nutshell

The definition:

> *"Social impact" or "making a difference" is (tentatively) about promoting total expected wellbeing — considered impartially, over the long term — without sacrificing anything that might be of comparable moral importance.*

Why "promoting"? When people say they want to "make a difference," we think they're primarily talking about making the world better — i.e. 'promoting' good things and preventing bad ones — rather than merely not doing unethical actions (e.g. stealing) or being virtuous in some other way.

Why "wellbeing"? We understand wellbeing as an inclusive notion, meaning anything that makes people better off. We take this to encompass at least promoting happiness, health, and the ability for people to live the life they want. We chose this as the focus because most people agree these things matter, but there are often large differences in how much different actions improve these outcomes.

Why do we say "expected" wellbeing? We can never know with certainty the effects that our actions will have on wellbeing. The best we can do is try to weigh the benefits of different actions by their probability — i.e. compare based on 'expected value.' Note that while the action with the highest expected value is best in principle, that doesn't imply that the best way to find the best action is to make explicit quantitative estimates. It's often better in practice to use rules of thumb, our intuition, or other methods, since these maximise expected value better than explicit expected value calculations. (Read more on expected value.)

Why "considered impartially"? We mean that we strive to treat equal effects on different beings' welfare as equally morally important, no matter who they are — including people who live far away or in the future. In addition, we think that the interests of many nonhuman animals should be given significant weight, although we're unsure of the exact amount. Thus, we don't think social impact is limited to promoting the welfare of any particular group we happen to be *partial* to (such as people who are alive today, or human beings as a species), but is about promoting welfare *impartially*.

Why do we say "over the long term"? We think that if you take an impartial perspective, then the welfare of those who live in the future matters. Because there could be many more future generations than those alive today, our effects on them could be of great moral importance. We thus try to always consider not just the direct and short-term effects of actions, but also any indirect effects that might occur in the far future.

Why do we add "without sacrificing anything that might be of comparable moral importance"? We aren't sure that improving welfare is the only thing that matters morally — moral philosophers have been arguing over what matters for a long time and it seems arrogant to assume we know the answer. Thus we think it's important to respect other values as well — e.g. autonomy and fairness. We find that this rarely comes up — respecting people's autonomy and promoting their welfare generally go hand in hand — but if there were a conflict, we would try very hard to avoid any actions that seem seriously wrong from one of these other common-sense perspectives.

Social impact is about making the world better

What does it mean to act ethically? Moral philosophers have debated this question for millenia, and have arrived at three main kinds of answers:

1. Making the world better — e.g. helping others.
2. Acting rightly — e.g. respecting the rights of others and not doing wrong.
3. Being virtuous — e.g. being honest, kind, and wise.

These correspond to consequentialism, deontology, and virtue ethics, respectively.

We think all three perspectives have something to offer, but when our readers talk about wanting to "make a difference," they're most interested in the first of these perspectives — changing the world for the better.

We agree this focus makes sense — we don't just want to avoid doing wrong, or live honest lives, but actually leave the world better than we found it.

And there is a lot we can all do to get better at that.

For instance, we've shown that by donating 10% of their income to highly effective charities, most college graduates can save the lives of over 40 people over their lifetimes with a relatively minor sacrifice.

From an ethical perspective, whether you save 40 lives or not will probably be one of the most significant questions you'll ever face.

In our essay on your most important decision, we argued that some career paths open to you will do hundreds of times more to make the world a better place than others. So it seems really important to figure out what those paths are.

In contrast, it's often a lot easier to know whether a path violates someone's rights or involves virtuous behaviour (most career paths seem pretty OK on those fronts), so there's less to gain from focusing there.

In fact, even people who emphasise moral rules and virtue agree that if you can make others better off, that's a good thing to do, and that it's even better to make more people better off than fewer. (And in general we think deontologists and utilitarians agree a lot more than people think.)

John Rawls was one of the most influential (non-consequentialist) philosophers of the 20th century, and said:

> *All ethical doctrines worth our attention take consequences into account in judging rightness. One which did not would simply be irrational, crazy.*[7]

Since there seem to be big opportunities to make people better off, and some seem to be better than others, we should focus on finding those.

So, while we think it's really important to avoid harming others and to strive to act virtuously, when it comes to real decisions, we think the potential positive consequences are what we should focus on the most.

This might sound like common sense, but it turns out to be an unusual way to look at things.

Much existing advice on 'ethical careers' is about *avoiding* working at bad companies rather than how we can do more good.

And discussion of ethical living more broadly typically focuses on reducing harm, rather than how we can best do good.

For instance, when it comes to fighting climate change, there's a lot of focus on our personal carbon emissions, rather than figuring out what we can do to *best* fight climate change.

Asking the second question suggests radically different actions. The *best* things we can do to fight climate change probably involve working on, advocating for, and donating

[7] See *A Theory of Justice* by John Rawls.

to exceptional research and advocacy opportunities, rather than worrying about plastic bags, recycling, or turning out the lights.

Why is there so much focus on our personal emissions? One explanation is that common-sense ethical thinking hasn't caught up with the situation that modern technology has put us in.

Our ethical views originate from before the 20th century, and sometimes from thousands of years ago. If you were a mediaeval peasant, your main ethical priority was to help your family survive, without cheating or harming your neighbours. You didn't have the knowledge, power, or time to help hundreds of people or affect the long-term future.

The Industrial Revolution gave us wealth and technology not even available to kings and queens in previous centuries. Now, many ordinary citizens of rich countries have enormous power to do good, and this means the potential consequences of our actions are usually what's most ethically significant about them.

So, we think 'social impact' or 'making a difference' should be about making the world better. But what does that mean?

What does it mean to make the world better?

We imagine building a world in which the most beings can have the best possible lives in the long term — lives that are free from suffering and injustice, and full of happiness, adventure, connection, and meaning.

There are two key components to this vision — impartiality and a focus on wellbeing — which we'll now unpack.

Impartiality: everyone matters equally

When it comes to 'making a difference,' we think we should strive to be *impartial* — i.e. to give equal weight to everyone's interests

This means striving to avoid privileging the interests of anyone based on arbitrary factors such as their race, gender, or nationality, as well as where or even *when* they live. In addition, we think that the interests of many nonhuman animals should be given significant weight, although we're unsure of the exact amount.

The idea of impartiality is common in many ethical traditions, and is closely related to the "Golden Rule" of treating others as you'd like to be treated, no matter who they are.

Acting impartially is an ideal, and it's not all that matters. As individuals, we all have other personal goals, such as caring for our friends and family, carrying out our personal projects, and having our own lives go well. Even considering only moral goals, it's plausible we have other values or ethical commitments beyond impartially helping others.

We're not saying you should abandon these other goals, and strive to treat everyone equally in all circumstances.

Rather, the claim is that *insofar* as your goal is to 'make a difference' or 'have a social impact,' we don't see good reason to privilege any one group over another — and that you should therefore have *some* concern for the interests of strangers and other neglected groups.

(And even if you think that the ultimate ideal is to have equal concern for all beings, as a matter of psychology, you probably have other, competing goals, and it's not helpful to pretend you don't.)

In Peter Singer's essay, *Famine, Affluence, and Morality*, he imagines you're walking and come across a child drowning in a pond. Everyone agrees that you should run in and save the child, even if it would ruin your new suit and shoes.

This illustrates a principle that many people can get behind: if you can help a stranger a great deal with little cost to yourself, that's a good thing to do. This shows that most people give *some* weight to the interests of others.

If it also turns out that you have a lot of power to help others (as we argued above), then it would imply that social impact should be one of the main focuses of your life.

Impartiality also implies that you should think carefully about who you can help the most. It's common to say that "charity begins at home," but if everyone's interests matter equally, and you can help more people who are living far away (e.g. because they're without cheap basic necessities you can provide), then you should help the more distant people.

We're convinced that a degree of impartiality is reasonable, but there remains huge questions about *how* impartial to be.

The trend over history seems to have been towards a greater impartiality and a wider and wider circle of concern, but we're unsure where that should stop. For instance, compared to people today, how should we weigh the interests of animals, people who don't exist yet, or even digital agents? This is called the question of moral patienthood.

Here's an example of the stakes of this question: we don't see much reason to discount the interests of future generations simply because they're distant from us in time. But because there could be so many people in the future, the main focus of efforts to do good should be to leave the best possible world for those future generations. This idea has been called 'longtermism,' and is explored in a separate article in the key ideas series. We think longtermism is an important perspective, which is why we say "over the long term" in our definition of social impact.

This section was about *who* to help; the next section is about *what* helps.

Wellbeing: what does it mean to help others?

When aiming to help others, our tentative hypothesis is that we should aim to increase their wellbeing as much as possible — i.e. enable more individuals to live flourishing lives that are healthy, happy, fulfilled; are in line with their wishes; and are free from avoidable suffering.

Although people disagree over whether wellbeing is the only thing that matters morally, almost everyone agrees that things like health and happiness matter a lot — and so we think it should be a central focus in efforts to make a difference.

Putting impartiality and a focus on wellbeing together means that, roughly, how much positive difference an action makes depends on how it increases the wellbeing of those affected, and how many are helped — no matter when or where they live.

What wellbeing consists of more precisely is a controversial question, which you can read about in this introduction by Fin Moorhouse. In brief, there are three main views:

- The hedonic view: wellbeing consists in your degree of positive vs negative mental states such as happiness, meaning, discovery, excitement, connection, and equanimity.
- The preference satisfaction view: wellbeing consists in your desires being fulfilled.
- Objective list theories: wellbeing consists in achieving certain goods, like friendship, knowledge, and health.

In philosophical thought experiments, these different views have very different implications. For instance, if you support the hedonic view, you'd need to accept that being secretly placed into a virtual reality machine that generates amazing experiences is better for you than staying in the real world. If instead you support (or just have some degree of belief in) the preference satisfaction view, you don't have to accept this implication, because your desires can include not being deceived and achieving things in the real world.

In practical situations, however, we rarely find that different views of wellbeing drive different decisions, such as about which global problems to focus on. The three notions correlate closely enough that differences in views are usually driven by other factors (such as the question of where to draw the boundaries of the expanding circle discussed in the previous section).

What else might matter besides wellbeing? There are many candidates, which is why we say promoting wellbeing is only a "tentative" hypothesis.

Preserving the environment enables the planet to support more beings with greater wellbeing in the long term, and so is also good from the perspective of promoting wellbeing. However, some believe that we should preserve the environment even if it doesn't make life better for any sentient beings, showing they place *intrinsic* value on preserving the environment.

Others think we should place intrinsic value on autonomy, fairness, knowledge, and many other values.

Fortunately, promoting these other values often goes hand in hand with promoting wellbeing, and there are often common goals that people with many values can share, such as avoiding existential risks. So again, we believe that the weight people put on these different values has less effect on what to do than often supposed, although they can lead to differences in emphasis.

We're not going to be able to settle the question of defining everything that's of moral value in this article, but we think that promoting wellbeing is a good starting point — it captures much of what matters and is a goal that almost everyone can get behind.

How good is it to create a happy person?

We've mostly spoken above as if we're dealing with potential effects on a fixed population, but some decisions could result in more people existing in the long term (e.g. deflecting an asteroid), while others mainly benefit people who already exist (e.g. treating people with parasitic worms, which rarely kill people but cause a lot of suffering). So we need to compare the value of increasing the number of people with positive wellbeing with benefiting those who already exist.

This question is studied by the field of 'population ethics' and is an especially new and unsettled area of philosophy.

We won't try to summarise this huge topic here, but our take is that the most plausible view is that we should maximise total wellbeing — i.e. the number of people (again including all beings whose lives matter morally) who exist in all of history, weighted by their level of wellbeing. This is why we say "total" wellbeing in the definition.

That said, there are some powerful responses to this position, which we briefly sketch out in the article on longtermism. For this reason, we're not certain of this 'totalist' view, and so put some weight on other perspectives.

Expected value: acting under uncertainty

How do you know what will increase wellbeing the most? In short, you don't.

You have to weigh up the different likelihoods of different outcomes, and act even though you're uncertain. We believe the theoretical ideal here is to take the action with the greatest **expected value** compared to the counterfactual. This means taking into account both how much wellbeing our actions could result in, and how likely those outcomes are, and adding them together.

In practice, we try to approximate this with rules of thumb, like the importance, neglectedness, tractability framework.

Going into expected value theory would take us too far afield, so if you want to learn more, check out our separate articles:

- Expected value: how to act when you're uncertain what will help?
- Counterfactuals and how they change our view of what does good
- How much risk to take if you want to do good?
- Cluelessness: can we know the effects of our actions?

Why do we say "without sacrificing anything that might be of comparable moral importance"?

In our definition, we say social impact is about promoting wellbeing, but we also add "without sacrificing anything that might be of comparable moral importance."

The purpose of this clause is to remind us of how much could be left out of our definition, and how radical our uncertainty is.

Many moral views that were widely held in the past are regarded as flawed or even abhorrent today. This suggests we should expect our own moral views to be flawed in ways that are difficult for us to recognise.

There is still significant moral disagreement within society, among contemporary moral philosophers, and, indeed, within our own team.

And past projects aiming to pursue an abstract ethical ideal to the exclusion of all else have often ended badly.

The “without sacrificing anything that might be of comparable moral importance” clause is a shorthand for the idea that we should:

- Consider a range of ethical perspectives, and take actions that seem good based on many perspectives
- Respect others’ values and be willing to compromise with them
- Respect our other important personal priorities, like family, personal projects, and our own wellbeing
- Be very cautious about doing anything that seems obviously wrong according to common sense

We think everyone has reason to be “morally uncertain” and so support these principles. You can read more about this in our separate articles:

- Is it ever okay to take a harmful job in order to do more good?
- Moral uncertainty: how to act when you’re uncertain about what’s good
- Coordination: why to respect other people’s values (even if you think they’re wrong)
- Ways people trying to do good accidentally make things worse, and how to avoid them

Is this just utilitarianism?

No. Utilitarianism claims that you’re morally obligated to take the action that does the most to increase wellbeing, as understood according to the hedonic view.

Our definition shares an emphasis on wellbeing and impartiality, but we depart from utilitarianism in that:

- We don’t make strong claims about what’s morally obligated. Mainly, we believe that helping more people is better than helping fewer. If we were to make a claim about what we *ought* to do, it would be that we should help others when we can benefit them a lot with little cost to ourselves, which is much weaker than utilitarianism.
- Our view is compatible with also putting weight on other notions of wellbeing, other moral values (e.g. autonomy), and other moral principles. In particular, we don’t endorse harming others for the greater good.
- We’re very uncertain about the correct moral theory and try to put weight on multiple perspectives.

Read more about how effective altruism is different from utilitarianism.

Overall, many of the team don't identify as being straightforward utilitarians or consequentialists.

Our main position isn't that people should be more utilitarian, but that they should pay more attention to consequences — and especially to the large differences in the scale of the consequences of different actions — than they do.

If one career path might save hundreds of lives, and another won't, we should all be able to agree that matters.

In short, we think ethics should be more scope sensitive.

Conclusion

We're not sure what it means to make a difference, but we think our definition is a reasonable starting point that many people should be able to get behind:

> *"Social impact" or "making a difference" is (tentatively) about promoting expected wellbeing — considered impartially, over the long term — without sacrificing anything that might be of comparable moral importance.*

We've also gestured at how social impact might fit with your personal priorities and what else matters ethically, as well as many of our uncertainties about the definition — which can have a big effect on where to focus.

We think one of the biggest questions is whether to accept longtermism, so we've dedicated the next article to that question.

From there, you can start to explore which global problems are most pressing based on whatever definition of social impact you think is correct.

You'll most likely find that the question of which global problems to focus on is more driven by empirical or methodological uncertainties than moral ones. But if you find a moral question is crucial, you can come back and explore the further reading below.

In short, if you have the extraordinary privilege to be a college graduate in a rich country and to have options for how to spend your career, it's plausible that social impact, as defined in this way, should be one of your main priorities. How to act on that priority effectively is the focus of the rest of the key ideas series.

Further reading on how to define social impact

- Podcast: Prof Will MacAskill on moral uncertainty, utilitarianism, and how to avoid being a moral monster

- A thread on why deontologists and utilitarians agree more than people think
- Radical Empathy by Holden Karnofsky
- *The Expanding Circle: Ethics and Sociobiology* by Peter Singer
- What is wellbeing? by Fin Moorhouse
- Impartiality on the Stanford Encyclopedia of Philosophy
- A reading list on moral patienthood on the Effective Altruism Forum

Longtermism: the moral significance of future generations

Click to read online

Most people think we should have some concern for future generations, but this obvious sounding idea can lead to a surprising conclusion.

Since the future might be very long, there could be far more people in future generations than in the present generation. This means that if you want to make the world better in an *impartial* way — i.e. without regard to people's race, class, or where or *when* they're born — then what most matters morally is that the future goes as well as it can for all generations to come. We've called this the 'long-term value thesis.'

When this thesis is also combined with the idea that some of our actions can have non-negligible effects on how the future goes, it implies that one of our biggest priorities should be ensuring the future goes well. This further idea is usually called 'longtermism.' Most of this article is about the long-term value thesis, but it comes back to whether we can affect the future at the end.

The long-term value thesis is often confused with the claim that we shouldn't *do* anything to help people in the present generation. But the thesis is about what most *matters* — what we should *do* about it is a further question. If it turned out that the best way to help those in the future is to improve the lives of people in the present, such as through providing health and education, then longtermists would focus on that. The difference is that the biggest *reason* to help those in the present would be to improve the long term.

The arguments for and against longtermism are a fascinating new area of research. Many of the key advances have been made by philosophers who have spent time in Oxford, like Derek Parfit, Nick Bostrom, Nick Beckstead, Hilary Greaves, and Toby Ord. We've found it interesting to watch them deepen and refine these arguments over the last 10 or so years, and we think longtermism might well turn out to be one of the most important discoveries of effective altruism so far.

In the rest of this article, you can see a one-page introduction to longtermism in a nutshell, then we give an overview of the main arguments for and against the long-term value thesis, discuss three common objections, and finish by briefly discussing whether we can affect the future and what longtermism might imply.

If you 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 send you a free copy when you subscribe to our newsletter!

If you prefer to start with a video, this presentation by Joseph Carlsmith is also a good introduction.

Longtermism in a nutshell

The average mammalian species lasts for about one million years. Homo sapiens have been around for only 200,000. With the benefit of technology and foresight, civilisation could, in principle, survive for at least as long as the earth is habitable — probably around 600–800 million years more.[8]

Given that we can't rule out this possibility, this means that there will, in expectation, be a huge number of future generations. There could also be a much larger number of people in each future generation, and their lives could be much better than ours.

We think future generations clearly matter, and impartial concern most likely implies their interests matter as much as anyone's.[9]

If we care about *all* the consequences of our actions, then what's most important about our actions from an impartial perspective is their potential effects on these future generations.

If this reasoning is correct, it would imply that approaches to improving the world should be evaluated mainly in terms of their potential long-term impact, over thousands, millions, or even billions of years.

In other words, the question "How can I have a positive impact?" should mostly be replaced with "How can I best make the very long-term future go well?" These

---

[8] And if humanity finds a way to flourish outside of Earth, civilisation could last far longer still.
[9] Technically, another claim is needed to complete this argument. We must say that others' interests matter regardless of their 'modal status' — roughly whether an individual would exist 'no matter what we do' or 'might exist or might not exist depending on our actions.' People who believe that this is an important moral difference often hold a 'person-affecting view.' We've given some reasons we disagree with the person-affecting view in our article on future generations.

arguments and their implications are studied as part of an emerging school of thought called *longtermism*.

We feel relatively confident about this idea, but we're not confident about what it implies in practice.

An obvious response to the above is that it's so difficult to predict the very long-term effects of our actions that although these effects might be very important, we don't know what they are. Instead, this response goes, we should focus on helping people in the short-term where we can have more confidence in their positive effects.

We agree it's very hard to know the long-term effects of our actions; however, as discussed above, we think we should aim to use whatever evidence and theory is available to make the best possible estimates of the expected value of different actions. Moreover, because the expected number of future generations is so great, our actions only need to have non-negligible effects on them for these effects to dominate their expected value.

In practice, we think there *are* some actions that potentially have very long-term positive effects. For example, we can take steps to make it less likely that civilisation ends through a disaster like nuclear war, which would irreversibly deprive future generations of the chance to flourish. We cover other examples in the next section.

Let's explore some hypothetical numbers to illustrate the general concept. If there's a 5% chance that civilisation lasts for ten million years, then in expectation, there are over 5,000 future generations. If thousands of people making a concerted effort could, with a 55% probability, reduce the risk of premature extinction by 1 percentage point, then these efforts would in expectation save 28 future generations. If each generation contains ten billion people, that would be 280 billion additional individuals who get to live flourishing lives. If there's a chance civilisation lasts longer than ten million years, or that there are more than ten billion people in each future generation, then the argument is strengthened even further.

Even if we're not sure what actions would help today, longtermism would likely imply that our key focus should be on carrying out research to identify these actions, or otherwise making society better able to tackle long-term challenges.

In contrast, we don't see much reason to expect that actions with good short-term effects will also be those that will be best from a long-term perspective.

Another major reason why we think it's worth taking longtermism seriously is that future generations lack any economic or political power, which means we should expect their interests to be neglected by our current institutions. Within philanthropy, too, very little attention is paid to the interests of those who might live more than 100 years in the future. This all suggests that outstanding opportunities to help may remain untaken, and that it would be reasonable for society to allocate significantly more attention to these issues.

We remain unsure about many of these arguments, but overall we're persuaded that focusing more on the very long-term effects of our actions is our best bet for now.

Why think the future matters more than the present?

What are the things you most value in human civilisation today? People being happy and free of suffering? People fulfilling their potential? Knowledge? Art?

In almost all of these cases, there's potentially a lot more of it to come in the future:

1. The Earth could remain habitable for 600–800 million years,[10] so there could be about 21 million future generations,[11] and if we do the needed work, they could lead great lives — whatever you think 'great' consists of. Even if you don't think future generations matter as much as the present generation, since there could be so many of them, they could still be our key concern.
2. Civilisation could also eventually reach other planets — there are 100 billion planets in the Milky Way alone.[12] So, even if there's only a small chance of this happening, there could also be dramatically more people *per generation* than there are today. By reaching other planets, civilisation could also last even longer than if we stay on the Earth.
3. If you think it's good for people to live happier and more flourishing lives, there's a possibility that technology and social progress will let people have much better and longer lives in the future (including people in the present generation). So, putting these first three points together, there could be many more generations, with far more people, with the potential to be living much better lives. The three dimensions multiply together to give the potential scale of the future.
4. It seems possible that the future could contain far more of the other things humans value, including beauty, justice, and knowledge.[13]
5. If you greatly value artistic and intellectual achievement, a far wealthier and bigger civilisation could have far greater achievements than our own.

---

[10] 800 million years from now: "Carbon dioxide levels fall to the point at which C4 photosynthesis is no longer possible. Free oxygen and ozone disappear from the atmosphere. Multicellular life dies out." Archived link, retrieved 22-October-2017.
[11] Assuming 3 generations per 100 years.
[12] *The Milky Way Contains at Least 100 Billion Planets According to Survey*, Hubblesite, Archived link, retrieved 22-October-2017
[13] Though some forms of justice and virtue focused ethics might not hold that we ought to *maximise* justice or virtue; instead, for instance, it may be a matter of satisfying a set of conditions.

And so on.

This suggests that, insofar as you care about making the world a better place, your key concern should be whether the future goes well or badly.

This isn't to deny that you have special obligations to your friends and family, and an interest in your own life going well. We're only talking about what matters insofar as you care about helping others impartially — philosophers often call this what matters "from the point of view of the universe." We think everyone should care about the lives of all others in this sense *to some degree*, even if you care about other things as well.

People often assume the long-term value thesis is especially about the possibility of there being lots of *people* in the future, and so only of interest to a narrow range of ethical views (especially total utilitarianism), but as we can see in the list above, it's actually much broader. It just rests on the idea that if *something* is of value, it's better to have more of what's valuable rather than less, and that it's possible to have much more of it in the future. This might include non-welfare values, such as beauty or knowledge. The arguments are also not about humans; rather, they concern whatever agents in the future might have moral value, including other species.

People also often think that the long-term value thesis assumes the future will have positive rather than negative value. Quite the opposite is true — the future could also contain far more suffering than the present, and this implies *even more* concern for how it unfolds. It's important to reduce the probability of bad futures as well as increase the probability of good ones.

You can see a more rigorous presentation of the arguments in Chapter 3 of *On the Overwhelming Value of Shaping the Far Future*, by Nick Beckstead.

Now let's consider three of the most common objections to the long-term value thesis.

1. Will the future actually be big?

The argument relies on the possibility of there being much more value in the future. But you might doubt that civilisation can actually survive very long, or that we will ever live on other planets, or that people's lives can be much better or worse than those of people today.

There's a lot of reason to doubt these claims. Let's look at them in a little more depth.

First, what's *not* up for debate is the *possibility* that the future could be big. It's a widely accepted scientific position that the Earth could remain habitable for hundreds of millions of years, and that there are at least a hundred billion planets in the galaxy. What's more, there's no reason to think it's impossible that civilisation could discover far more powerful technology than we have today, or that people could live far better and more satisfying lives than they do today.

Rather, what's in doubt is the *likelihood* that these developments come to pass. Unfortunately there is no definitive way to estimate this likelihood. The best we can do is to weigh the arguments for and against a big future, and make our best estimates.

If you think that civilisation is virtually guaranteed to end in the next couple of hundred years, then the future won't have much more value than the present. However, if you think there's a 5% chance that civilisation survives 10 million generations till the end of the Earth,[14] then (in expectation) there will be over 500,000 future generations. This means the future is at least 500,000 times 'bigger' than the present. This could happen if there's a chance civilisation reaches a stable state where the risk of extinction becomes low.

In general, the bigger you think the future could be, and the more likely you think we are to get there, the greater the value.

Further, if you're *uncertain* whether the future will be big, then a top priority should be to *figure out* whether it will be — it would be the most important moral discovery you could make. So, even if you're not sure you directly should act on the thesis, it might still be the most important area for research. We see this kind of research as extremely important.

2. What about discounting?

Sometimes at this point people, especially those trained in economics, mention "discounting" as a reason to not care about the long term.

When economists compare benefits in the future to benefits in the present, they typically reduce the value of the future benefits by some amount called the 'discount factor.' A typical social discount rate might be 1% per year, which means that benefits in 100 years are only worth 36% as much as benefits today, and benefits in 1,000 years are worth almost nothing.

To understand whether this is a valid response, you need to consider why the concept of discounted benefits was invented in the first place.

There are good reasons to discount *economic* benefits. One reason is that if you receive money now, you can invest it, and earn a return each year. This means it's better to receive money now rather than later. People in the future might also be wealthier, which means that money is less valuable to them.

[14] Sometimes people object that this is a "Pascal's wager" type argument, but a 10% chance is typically larger than is used in these arguments, and the supposed future value is still finite rather than infinite.

However, these reasons don't obviously apply to welfare — people having good lives. You can't directly "invest" welfare today and get more welfare later, like you can with money. The same seems true for other intrinsic values, such as justice.

There are other reasons to discount welfare,[15] and this is a complex debate. However, the bottom line is that almost every philosopher who has worked on the issue doesn't think we should discount the *intrinsic value* of welfare — i.e. from the point of view of the universe, one person's happiness is worth just the same amount no matter when it occurs.

Indeed, if you suppose we can discount welfare, we can easily end up with conclusions that sound absurd. For instance, a 3% discount rate would imply that the suffering of one person today was equal to the suffering of 16 trillion people in 1,000 years.

As Derek Parfit said:

> *Why should costs and benefits receive less weight, simply because they are further in the future? When the future comes, these benefits and costs will be no less real. Imagine finding out that you, having just reached your twenty-first birthday, must soon die of cancer because one evening Cleopatra wanted an extra helping of dessert. How could this be justified?*

If we reject the discounting of welfare and other intrinsic values, then the chance that there could be a great deal of value in the future is still important. Moreover, this doesn't stand in tension with the economic practice of discounting monetary benefits.

If you'd like to see a more technical discussion of these issues, see *Discounting for Climate Change* by Hilary Graves. There is a more accessible discussion at 1h00m50s in our podcast with Toby Ord and in Chapter 4 of *Stubborn Attachments* by Tyler Cowen.

3. Do we have moral obligations to future generations?

A final response is that although the future might be big, we're not obligated to help people who don't yet exist in the same way as we're obligated to help people alive right now.

This objection is usually associated with a "person-affecting" view of ethics, which is sometimes summed up as the view that "Ethics is about helping make people happy, not making happy people." In other words, we only have moral obligations to help those who are already alive, and not to enable more people to exist with good lives.

[15] For instance, we might discount welfare due to uncertainty about whether it will happen, but we've already taken uncertainty about the future into account when estimating its expected value.
We might also discount welfare in practice, since having happy people now might be thought to produce more happy people in the future.

You can see where this intuition comes from if you consider the following choice: is it better to cure one person who's 60 years old of cancer and allow them to live to 80, or to bring one new person into existence who will live a good life for 80 years? Most people think we should help the 60 year old, even though the new person gains four times as much good life.

However, person-affecting views suffer from a number of problems. For instance, suppose you have the option to bring into existence someone who would not otherwise have existed, whose life involves severe and constant suffering from birth until death, and who wished they had never been born.

Nearly everyone agrees this is a bad thing to do. A naive person-affecting view, however, says that, since it involves creating a new person, this lies outside of our ethical concern, and so is neither good nor bad. So, the person-affecting view conflicts with the obvious idea that we shouldn't create the suffering life.

Person-affecting views can avoid this conflict by positing that it's bad to create lives filled with suffering, but it's neither good nor bad to create *happy* lives. Then, it's wrong to create the suffering-filled life, but there's no reason to enable more happy people to exist in the future.

One issue with this is that it's unclear why this asymmetry would exist. The bigger problem though is that this asymmetry conflicts with another common sense idea.

Suppose you have the choice to bring into existence one person with an amazing life, or another person whose life is barely worth living, but still more good than bad. Clearly, it seems better to bring about the amazing life, but if creating a happy life is neither good nor bad, then we have to conclude that both options are neither good nor bad. This implies both options are equally good, which seems bizarre.

This is a complex debate, and rejecting the person-affecting view also has counterintuitive conclusions. For instance, if you agree that it's good to create people whose lives are more good than bad, then you'll need to accept that we could have a better world filled with a huge number of people whose lives are just barely worth living. This is called the "repugnant conclusion."

Our view, however, is that it's better to reject the person-affecting view. You can see a summary of the arguments in this public lecture by Hilary Greaves (based on this paper) and in Chapter 4 of *On the Overwhelming Importance of Shaping the Far Future* by Nick Beckstead. It's also discussed in our podcast with Toby Ord.[16]

[16] It's also been argued that even if you take a person-affecting approach, you might still think that the future is more important than the present. This is because some people who are alive today might be able to live a very long time, and have much higher levels of welfare than they do today. Rather than

What's our position? As stated, we find the criticisms of the person-affecting view persuasive, so don't find it a convincing response to the long-term value thesis. However, since many people hold something like the person-affecting view, we think it deserves some weight, and that means we should act as if we have somewhat greater obligations to help someone who's already alive compared to someone who doesn't exist yet. (This is an application of moral uncertainty).

Likewise, we think there might be other types of ethical reasons to have additional concern for the present. For instance, maybe the unique nature of injustice means we have *extra* reasons to fight great injustices being perpetrated today. (Though there may also be reasons of justice to make sure the interests of future generations aren't ignored.)

However, these reasons to place special value on the present need to be set against the potentially far greater amount of value in the future. How to do this is an extremely difficult question, and involves unsettled questions in the study of moral uncertainty.

Trying to weigh this up, we think we should have far greater concern for the future, though we care more about the present generation than we would if we naively weighed up the numbers.

So, is there much more value in the future?

We think the original argument survives the responses, and so, when it comes to doing good, our key concern is how the long-term unfolds.

That said, we're still highly uncertain about these arguments. There's a good chance we've missed a crucial consideration and this picture is wrong. These ideas are still new and have not been heavily studied. We're also uncertain how to weigh the value of the future against other moral concerns given moral uncertainty.

This makes us cautious to put *overwhelming* value on the future, even if that's what the raw numbers might imply. Instead, we see making the future go well as our key, but not only, moral concern.

We also place a great amount of value on learning more about these issues to refine our priorities, and we attach importance to many other moral demands.

Can we actually influence the future?

You might be persuaded by these arguments, but believe they are irrelevant because we can't significantly impact the future. It's natural to think that the overall effects of our

create more people, your key concern should be to ensure that these possibilities are realised. So, even if you have a person-affecting view, the long-term value thesis might still hold.

actions on the future are unknowable. If so, you might accept the long-term value thesis but reject longtermism, and instead believe that the best we can do is help people in the short term.

However, we think there are ways we can impact the future:

1. We can speed up processes that impact the future. Our economy tends to grow every year, and this suggests that if we make society wealthier today, this wealth will compound, making future people wealthier.
2. More significantly, we could precipitate the end of civilisation, perhaps through a nuclear war, run-away climate change, engineered pandemics, or other disasters. This would foreclose the possibility of all future value. It seems like there are things the current generation can do to increase or decrease the chance of extinction, as we cover in our problem profiles.
3. There might be other major, irreversible changes besides extinction that we can influence, which could either be positive or negative. For instance, if a totalitarian government came into power that couldn't be overthrown, the badness of that government would be locked-in for a very long-time. If we could decrease the chance of that happening, that would be very good, and vice versa. Alternatively, genetic engineering might make it possible to fundamentally change human values, and then these values would be passed down to every future generation. This could either be very good or very bad, depending on your moral views and how it was done.
4. Even if you're not sure how to help the future, then your key aim could be to do *research* to work it out. We're uncertain about lots of ways to help people, but that doesn't mean we shouldn't try. This is part of global priorities research, and there are plenty of concrete questions to investigate.

You can see more detail on the different ways to shape the future in Chapter 3 of *On the Overwhelming Importance of Shaping the Far Future*.

What's the practical upshot of these possibilities?

If you think there's a huge amount of value in the future, and there are not-ridiculously-small ways we can affect it, then these actions will be the highest-impact we can take. Nick Beckstead calls this the "future-shaping argument."

In fact, it turns out that many of the ways to help future generations are also highly *neglected*. This is exactly what you'd expect — the present generation has a much greater interest in helping itself rather than improving the future.

Future generations can't buy things, so lack economic power. They lack any political representation, and depend entirely on our generosity towards them. And because we never see the effects of our actions, even people who want to make a difference often neglect them. Within philanthropy, too, very little attention is paid to the interests of those who might live more than 100 years in the future.

This all suggests that outstanding opportunities to help may remain untaken, and that it would be reasonable for society to allocate significantly more attention to these issues.

You can read more about objections to longtermism and responses in this article.

What are the best ways to help future generations right now?

This is a topic for another article, but here is a very quick outline of our views.

One area that *doesn't* seem to be the top priority is speeding up progress. In terms of importance, Beckstead argues in Chapter 3 of his thesis (expanding on Nick Bostrom's original paper, Astronomical Waste) that from a long-term perspective, what matters most is where we end up, not how fast we get there, so speed-ups are less important than changes that alter the long-term trajectory.

Second, efforts to speed up progress are also far less neglected than other ways to help the future — the world spends about one trillion dollars a year on R&D, and a lack of neglect is what we should expect, since these discoveries also benefit the present generation.

Instead, what's most important are "path changes" — actions that have the potential to shape the future over a very long timescale. (You could argue that speeding up progress will result in a path change, but the path change is where most of the value is.)

The main question is then which types of path change we should focus on?

The clearest example of a priority today seems to be reducing existential risks.

There is a small but real possibility that civilisation ends in the next century; and this would not only be terrible for the present generation, it would permanently remove the possibility of a good future. We need to get these risks down before we can focus on other ways of improving the future. What's more, there are many concrete, highly neglected proposals to reduce these risks.

Read more about the arguments for focusing on existential risks, which apply even if you're mainly focused on the present generation, but even more so if you accept the long-term value thesis.

Not everyone focused on longtermism thinks we should focus on directly reducing specific existential risks. Some think we should try to shape emerging technologies to increase the chance they go well (as well as reduce risks), others think we should focus on making society generally better able to navigate challenges (since perhaps we don't know what the biggest risks are), and yet others think we should focus on building the resources of those concerned about longtermism in the future, so they're in a better position to act compared to us — this is called patient longtermism. We discuss these options more in a podcast with Benjamin Todd.

We're also deeply unsure about all of these suggestions, so our other focus is on global priorities research to identify the best ways to help the future. This includes the question of whether there might be positive path changes to promote (sometimes called "existential hope"), as well as what negative risks are most important to avoid.

There might also be crucial considerations that might overturn longtermism. We think this research could have a significant impact on our priorities, and there's also only a handful of researchers currently doing it.

The stakes facing our generation are much more than they first seem. Our actions might have the potential to bring about a far better world, or cut it short. Our key concern should be to ensure the future goes well.

Want to focus your career on the long-run future?

If you want to work on any issues essential to ensuring the future goes well, such as controlling nuclear weapons or shaping the development of artificial intelligence or biotechnology, you can speak to our team one-on-one.

We've helped hundreds of people choose an area to focus, make connections, and then find jobs and funding in these areas. If you're already in one of these areas, we can help you maximise your impact within it.

Further reading

- Moral uncertainty: how to act when you're uncertain about what's good
- Chapters 1-4 *On the Overwhelming Importance of Shaping the Far Future* by Nick Beckstead. This thesis builds upon a seminal paper by Nick Bostrom, *Astronomical Waste*, which makes the argument that the future has overwhelming value for utilitarians, as well as *Existential Risk Prevention as a Global Priority*.
- A more recent and systematic academic treatment of the topic can be found here: The Case for Strong Longtermism, by Hilary Greaves and Will MacAskill.
- Dr Nick Beckstead also discusses his thesis in our podcast.
- Dr Toby Ord discusses these arguments in his book, *The Precipice.*
- Dr Toby Ord also explores many of these arguments with us in another podcast.

- Benjamin Todd and Arden Koehler discuss varieties of longtermism in this podcast.
- Our podcast with Alexander Berger presents a case for focusing instead on near-term gains in global health and wellbeing.
- In this podcast, Holden Karnofsky talks about the case that we're living in the most important century. And in a blog post, *Are we living at the most influential time in history?*, Dr Will MacAskill responds with counterarguments.
- This could be the most important century.
- Why despite global progress, humanity is probably facing its most dangerous time ever.
- Fifty-one policy and research ideas for reducing existential risk.
- Carl Shulman on the common-sense case for existential risk work and its practical implications.
- An alternative introduction to longtermism, which goes through somewhat different objections and reasoning.

Effective altruism

Click to read online

It sounds obvious that it's better to help two people than one if the cost is the same. However, when applied to the world today, this obvious-sounding idea leads to surprising conclusions.

In short, we believe we need a new approach to having an impact: effective altruism.

Modern levels of wealth and technology have given some members of the present generation potentially enormous abilities to help others, while our common-sense views of what it means to be a good person have not caught up with this change.

This means that some actions that are widely considered to 'do good' have dramatically greater positive consequences than others.

For instance, the UK's National Health Service and many US government agencies are willing to spend over $30,000 to give someone an extra year of healthy life.[17] This is a great use of resources by ordinary standards.

However, research by GiveWell has found that it's possible to give an infant a year of healthy life by donating around $100 to one of the most cost-effective global health

[17] "Below a most plausible ICER of £20,000 per QALY gained, the decision to recommend the use of a technology is normally based on the cost-effectiveness estimate and the acceptability of a technology as an effective use of NHS resources." From *Guide to the methods of technology appraisal 2013,* by the National Institute for Health and Care Excellence. Archived link

charities, such as the Against Malaria Foundation (AMF). This is about 0.33% as much. This suggests that at least in terms of improving health, one career working somewhere like AMF might achieve as much as 300 careers focused on typical ways of doing good in a rich country.

These kinds of health programmes offer such a good opportunity to do good that even the most prominent aid sceptics have offered few arguments against them.

This is not unique to health programmes. When we've looked at other ways of doing good, this pattern tends to be replicated: the most effective ways to help usually seem much better than what's typical. For instance:

- Among ways individuals can combat climate change, there are huge differences in how much different actions reduce CO2.
- Within US social programmes, the most cost effective are far more effective than the median.
- Within politics, the best get out the vote methods work far better than typical ones.

Indeed, many social programmes have little impact at all, which by itself creates a significant difference between the best and typical.

In our own work, we've argued that the impact of different career paths depends on:

- How pressing the problems you work on are
- How effective the solutions you support are
- How much leverage you can apply to those solutions
- Your degree of personal fit with the path

And in the rest of the key ideas series, we argued that the best vs. typical options within each factor often seem to vary by a factor of 10 and sometimes a lot more (i.e. the paths with the most leverage open to you have more than 10 times the leverage of a more typical option, the most effective solutions make 10 times more progress on the problem per year of work, etc.). And since the overall impact of a career is given by the multiple of the four factors, the total variation in impact is even larger.

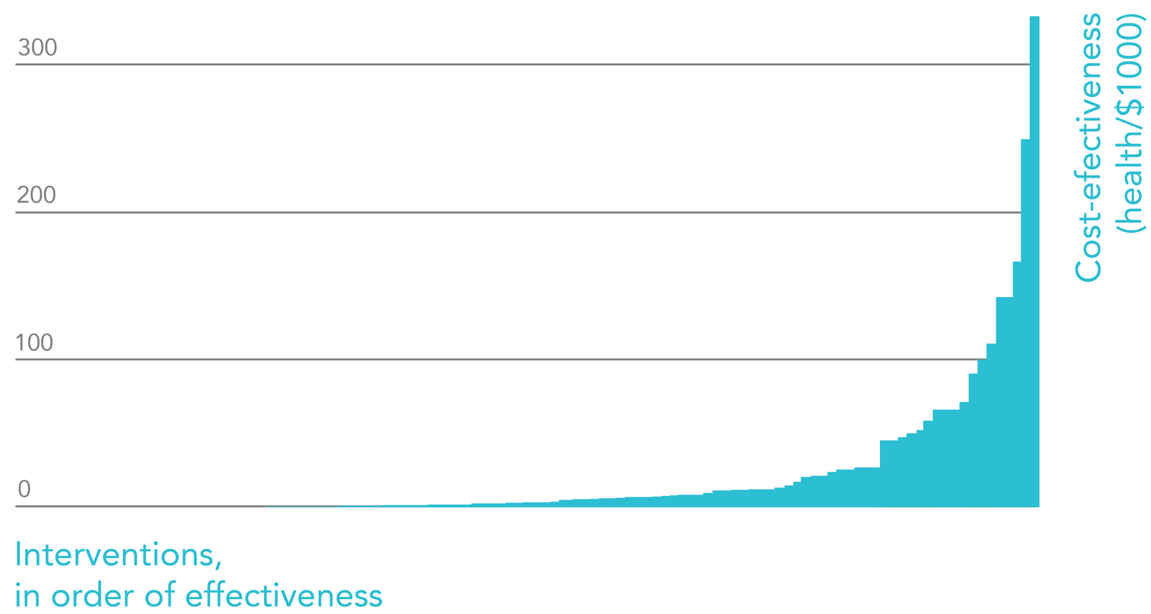
This wide spread of outcomes is also probably what we should expect to find.

A trait like height follows a 'normal' distribution: the tallest people are only about 50% taller than the average. A trait like income, however, follows a 'fat tail' distribution: the highest-earning people earn thousands of times more than average. This concept has been popularised as the '80/20 principle' — which says that 80% of outcomes are caused by 20% of events — or as the idea that outcomes are dominated by 'black swan events.'

We expect that the distribution of the expected impact of different actions is more likely to be like income than height.

One reason for this is that if the outcomes of different actions are caused by the multiplication of several factors — as they often are — then the value of different actions will end up as a heavy-tailed distribution (technically, a log-normal distribution).

This is the pattern we see where we have good data. For instance, the cost effectiveness of ways to improve health in poor countries (measured with 'cost per year of healthy life') closely follows an 80/20 distribution.



Research by Open Philanthropy finds a similar pattern in other areas.

This means that if your aim is to impartially help others, and all else equal you'd prefer to help more people, your key concern shouldn't just be to 'make a difference' — it should be to identify the *very best* ways to help among the options open to you.

This idea might sound obvious, but when we surveyed people on how much more effective they think the best charities are compared to the median, a typical response was that the best charities are only 66% more effective; whereas instead it seems like the difference is more like 10,000%. So, the difference between the best and typical ways of helping are much larger than ordinarily supposed.

This idea also has some radical implications.

One big implication is that it means that people who want to do good could achieve far more than they do today, simply by paying more attention to the effectiveness of their options.

For instance, if some ways of doing good achieve 100 times as much as others, then we could be making 100 times as much progress on global problems without using any more resources. Or to put it another way, it would be like increasing the number of people focused on doing good by 100 times.

As an individual, it really seems possible to save the lives of tens of others, or even to play a role in reducing the risk of the end of civilisation or shaping the most important century.

Perhaps an even bigger implication is that we need to rethink what it means to be a good person.

Our ethics are very focused on not doing wrong, rather than doing *more* good. This made sense in a pre-modern world where we didn't have much power to do good. But today it's missing the most important part of the picture.

This is clear in discussions of climate change, where most attention is given to how individuals can reduce their personal climate footprint. But if you try to reduce your personal climate footprint, the best-case scenario is avoiding around 5-20 tonnes of emissions per year (roughly the average emissions of the citizens of a rich country) — often at great expense.

Instead we should be asking, "What's the best thing I can do to tackle climate change?" By making well-targeted donations, through carefully targeted advocacy efforts, or using your career to tackle climate change, we think you could do vastly more than avoid 20 tonnes of emissions per year, and probably with less personal sacrifice.

Rather than not doing wrong, modern ethics should be focused far more on how we can best do good.

So how can we best do good? Despite the vital importance of this question, it's received very little direct study. We need a new field that applies our best methods for seeking truth to answer it.

And it turns out that while 'helping two people is better than one' and 'let's use reason and evidence to find the best ways to help' sound like obvious principles, if you follow them where they lead, you end up in some pretty weird places. It's perhaps similar to how using reason and evidence to search for truth led us to quantum mechanics.

This means we also need to rethink which actions actually do the most to help.

These are the insights behind the 'effective altruism' movement, which we helped to found in 2012. Effective altruism is the search for the best ways to help others. It can be divided into two projects:

- A research field aimed at finding the actions that achieve the most good.
- A community of people around the world who work together to put the ideas in practice and have a positive impact.

To think effective altruism is worthwhile in principle, all you need to believe is that it's good to help others, and that you'd prefer to help more people rather than fewer.

If we combine this with the opening claim that some ways of helping are much more effective than others, but they're not what people are focused on today, then it implies we need a new field of research to search for the best ways of helping, and that by applying the findings, we can do a lot more good.

We see 80,000 Hours as a project within the broader project of effective altruism. Our focus is on finding the career paths and strategies that enable you to best help others, but effective altruism in general is about all ways of doing good: through donations, career change, volunteering, consumption, campaigning, etc.

There's now a global community of thousands of people trying to put these ideas into practice, and tens of billions of dollars committed to the approach. Here's how to get involved.

Below, you can either learn more about effective altruism, or you can continue with the key ideas series to see how we apply effective altruism to career choice.

Learn more about effective altruism

You have a choice of introductions:

- Our podcast series: Effective Altruism: an introduction
- An online article: An Introduction to Effective Altruism
- An academic article: Effective Altruism by Will MacAskill

You can also read *Doing Good Better*, which is by our co-founder and Oxford philosopher, Will MacAskill. Steven Levitt, the author of *Freakonomics*, said the book "should be required reading for anyone interested in making the world better." The book is more focused on measurable global health interventions than we are today, but discusses the same principles. Will MacAskill also has a more recent TED talk.

Further reading

- What is the core idea of effective altruism? an interview with our founder, Benjamin Todd
- The guiding principles of effective altruism by the Centre for Effective Altruism
- Effective altruism by GiveWell
- Common objections to effective altruism
- Misconceptions about effective altruism

Other articles on foundations

- Expected value: how can we know what makes a difference when uncertain?
- Why good judgement is so crucial to doing good, and how to improve
- Is it ever OK to take a harmful career for the greater good?
- How to estimate very uncertain things. An introduction to practical bayesianism.
- Moral uncertainty: how to act when uncertain you're about what's good
- Cluelessness: can we ever know the effects of our actions?
- Counterfactuals and how they change our view of what does good

3. Global priorities: what's the world's most pressing problem?

Why the problem you work on is the biggest driver of your impact

Click to read online

If you want to make a difference, which issue is best to work on — climate change, education, pandemics, or something else?

People often think that making these kinds of comparisons is near impossible. The most common advice is that you should just work on whatever issue you're passionate about.

But we believe some global problems are far bigger and more neglected than others, and so which issues you work on will probably be the biggest driver of your impact.

An introduction to comparing global problems

We'd like to see a great many global problems get more attention, but as individuals, the best we can do is identify the biggest gaps in existing efforts and help fill them.

To find these gaps, one starting point is to look for problems that are:

1. Important: if progress is made, how much social impact would result?
2. Neglected: how much effort will be invested in this problem by others?
3. Tractable: how easy is it to make progress per unit of resources?

In this article, we'll argue that there are huge differences in how important and neglected different issues seem, which don't seem to be offset by differences in tractability.

This means that by choosing a different issue, you might be able to increase how much impact you have by over 100 times.

Some problems are bigger than others

Climate change is widely considered one of the world's biggest problems, and we think it's even bigger than often supposed. While the most likely scenario is several degrees of warming, the uncertainty in climate models means it's hard to rule out warming over 10°C by 2200.

What's more, the CO_2 we emit today will stay in the atmosphere for tens of thousands of years, impacting our children's grandchildren and beyond. We think future generations matter, which makes the issue even bigger in scale.

But we also think there may be issues that are even larger still.

The philosopher Toby Ord has argued that in 1945 humanity entered a new age, which he calls 'the Precipice.' On July 16, 1945, humanity detonated the first atomic bomb, which would eventually make it possible — for the first time in history — that a small group of people could destroy most of the world's cities within hours.

The annual risk of an all-out nuclear exchange is small, but it's not zero: there is always the chance of an accident or malfunction.

The average of several expert surveys estimated the chance of a US-Russia exchange is 0.4% per year.

The probability of an all-out exchange is lower, but over our lives and the lives of our children, it could still add up to a substantial chance of a catastrophe potentially more devastating than climate change.

Besides killing most people living in urban areas, the resulting fires could lift enough ash into the air to obscure the sun and reduce global temperatures for years, leading to widespread famine through a phenomenon known as 'nuclear winter.'

Within months, this would not only kill most people alive today, but it could also lead to a collapse of civilisation itself. We think a permanent collapse is very unlikely, but when we consider the scale of the consequences — the loss of all future generations — that risk may be the worst thing about a nuclear conflict.

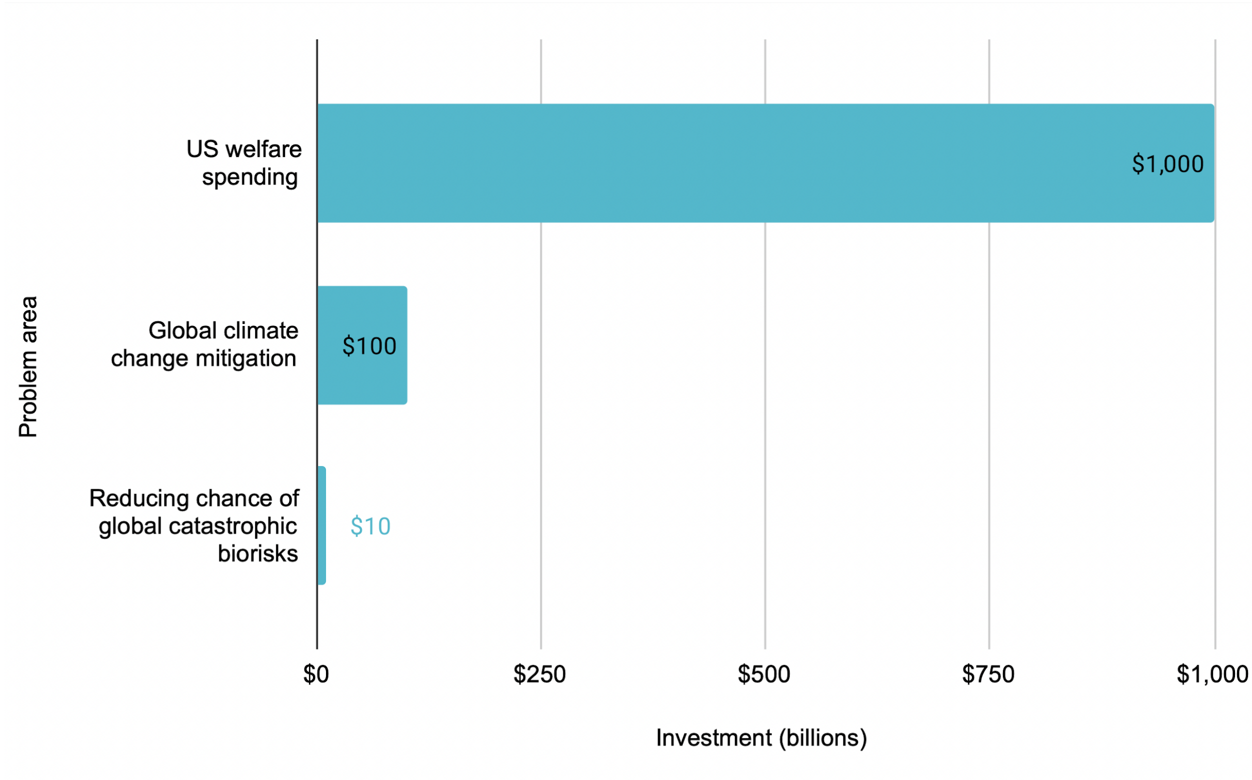
But the possibility of extreme climate change and nuclear winter are just two examples of a broader trend.

Technology has given this generation unprecedented power to shape history. The consequences of the decisions we make today about nuclear weapons, genetic engineering, artificial intelligence, space settlement, and other emerging technologies could ripple forward for thousands of years, with either enormously positive or enormously negative consequences.

So: some problems are *much bigger* than others.

Some problems are more neglected than others

In 2016 we argued that a global pandemic posed a significant global risk, but with around $10 billion spent globally on preventing the worst pandemics per year,[1] it was more neglected than climate change or international development (which receive hundreds of billions) — which are in turn far more neglected than education and health in rich countries (which receive trillions).



Why work on issues that are comparatively neglected? At least among issues that are roughly similar in importance, it's usually harder to have a big impact working on more established or popular issues, because there are probably already people working on the most promising interventions. For this reason, if you're the 100th person working on a problem, your contribution is likely to make a much larger difference than if you're the 10,000th.

How much larger? In our view, returns to more work diminish relatively quickly, and approximately 'logarithmically' — meaning that it matters a lot how neglected an area is.

From what we've seen, some global issues appear to be *thousands of times more neglected* than others of similar importance — they receive only a tiny fraction of the resources. This implies that if all else is held constant, work in some areas is *thousands of times more effective* than work in others.

Inspired by this argument, one of our readers, Cassidy, realised that her knowledge as a doctor might be put to even better use in pandemic prevention.

She applied to a masters in public health programme, and from there was able to get into a biosecurity PhD at Oxford. Since the field of pandemic control was relatively small before COVID-19, she was able to advance quickly. When COVID-19 broke out, she was ready to advise the UK government on policy ideas to control COVID-19, as well as how to prevent the next (potentially much worse) outbreak.

Cassidy realised she could help more people by working on pandemic policy than by seeing one patient at a time as a doctor.

The importance of working on neglected issues means that following your current passions could easily point you in the wrong direction. You're most likely to stumble across the same issues everyone is already talking about, which will usually be among the *least* neglected. The best options are probably unconventional.

Which issues might be even more neglected than pandemics?

A survey of AI researchers in 2017 found that they believe there's about a 50% chance that AI systems will exceed human capacities in most jobs around 2060. This would be one of the most important events in history.

These same researchers also estimated that if this happened, while the outcome might be 'extremely good,' there is also a 5% chance the outcome could be 'extremely bad' (e.g. human extinction). And as the people developing the technology, they're probably on the optimistic side.

One reason some people are concerned is that it's unclear whether AI systems will continue to stay aligned with human values as they become more powerful. This could suggest that research into 'AI alignment' is a crucial challenge from a long-term perspective.

Since we first wrote about it in 2012, AI alignment has grown into a flourishing field of research in computer science, but it still receives under $100 million in funding per year — about 100 times less than preventing pandemics.

Long-term AI *policy* — addressing the question of how society and government should handle advancing AI — is even more neglected. (Read more about our case for both.)

How much do problems differ overall?

How important and neglected a problem is multiply together to determine how pressing the issue is overall.

We've argued that some issues seem over 100 times bigger than others, and some seem over 100 times more neglected than others.

Since it's often the most important issues that seem the most neglected, this would suggest that overall you'll have over 10,000 times as much impact working on some issues rather than others (all else equal).

However, there are some strong counterarguments to acting as if the spread is this large, which we discuss in our podcast about the key ideas series. Overall, we think the differences are more like 1,000-fold.

For instance, based on our view of global priorities, we think by focusing on global poverty rather than a typical social issue in a rich country, you might have 10–100 times as much impact, and then by focusing on existential risk, you might have 10–100 times as much impact again.

You might differ from us in which issues you think are most pressing, but we expect you'll still conclude there are very large differences.

What does this spread imply?

In an ideal world, there would be far more people working on every important social issue. But each of us only has one career, and we'll all have far more impact if we focus on the issues that are the most pressing for us to work on.

If it's possible to have 100 or even 1,000 times as much impact per year by changing the issue we focus on, that's a huge deal. It would probably be the single biggest thing you could do to increase the impact of your career.

If you don't have the option of making a big career change right now, there's a lot you can do to support the most pressing issues no matter your current job, through donations, political engagement and mobilising others. But how about working on these issues directly?

Sometimes it's relatively easy to support a new issue in your existing role. For instance, if you work in media, you might be able to tilt which issues you cover.

Our readers are sometimes surprised at how their existing skills can be applied to unusual problems like AI alignment.

Brian Tse was working at an investment bank in Hong Kong and didn't have a background in AI. But he started to learn more about it in his spare time, and using his bilingual background, began translating materials to connect Western and Chinese researchers.

He now runs an independent consulting firm advising organisations in China, the United States, and Europe on the safety and governance of AI and other potentially transformative technologies — a path he also finds much more fulfilling and exciting than banking.

While working in the corporate world, Brian learned about AI and other transformative technologies. He now consults with many key Chinese and Western organisations that work on AI, with the aim of helping them coordinate better.

Many people working in government have significant flexibility about which areas of policy they work on. For instance, Clíodhna was working in health policy, but after learning more about AI alignment, she was able to switch to working at the intersection of AI and health.

But even if the switch seems hard, the huge potential gains could mean it's easily worth it — even if you'd need to retrain, take a more junior role, or test out a role where you're not sure you'd be a good fit or aren't sure how to make progress on the issue.

People we advise are often tempted to go for an issue they think is second-tier because it seems easier to enter. But if a top-tier issue might have 10 times the impact, it's often worth spending some time testing out your fit, even if you're not sure it'll work out.

That's not to downplay the difficulty — orienting your career around a new problem is a big decision. Our main message is that it deserves some very, very serious thought — and much more attention than it normally gets.

Choosing a problem and your career: some common misunderstandings

We discuss how choosing a problem fits into the rest of your career plan and strategy in more depth in our planning process, but here are a couple of quick clarifications.

Do I need to pick an issue right away?

No. The importance of choosing a problem means it's worth giving yourself time to reflect on your worldview and learn about different issues. People's views on these questions often develop over the course of several years, and that's time well spent.

Early in your career (say under 30), it's often more important to invest in career capital that can give you more leverage to work on whichever issues turn out to be most pressing in the future, when you're at your peak productivity. In the meantime, it's usually enough just to have a broad sense of the types of issues you want to work on, and can narrow down later.

One important exception is if you're considering making a big commitment to a particular issue, such as a specialised PhD, or a particular career track, such as medicine. In this case, we'd encourage you to think more about your choice of issue right away.

And if you're lucky enough to have already found a promising way to tackle a big problem early on, by all means get started right away.

If you're later in your career, then the most important decisions you face are probably which issues to focus on, and how to use your existing career capital to contribute to them.

Are you saying everyone should work on the top issue?

No. People often think we think everyone should work on AI, but this isn't the case.

We think the impact you have over your career depends on:

1. How pressing the problems you focus on are
2. How effective the solutions you pursue are
3. The amount of leverage you can apply to those solutions
4. Your personal fit for the path

Although we think the first factor is often the most important, the other three factors also matter a great deal. So, even if you agree with us about which problems are most pressing, it doesn't mean you should only work on the top one.

For instance, we once advised someone who was choosing between a senior position in international development policy, or starting at the bottom in emerging technology policy. It seemed like they would have about 100 times as much leverage in the senior position, and it was also a better personal fit, so we recommended international development over AI.

As another example, we often encourage people with biology backgrounds to work on pandemics over AI due to their better fit.

Coordination and why to spread out over issues

Considering coordination gives us yet more reasons to spread out over a range of problems.

Ultimately, our readers help form a community of thousands of people trying to tackle global problems, which broadly overlaps with the effective altruism community. This community should work on a 'portfolio' of issues for several reasons:

1. Diminishing returns: there are often a couple of especially good but time-sensitive opportunities within each issue each year. It's important to have at least some people working on each issue so they're able to take these.

2. Information value: by working on an issue, you learn more about how effective it is. While we think most people should work on the top issues, a minority should spread out over a wide range of plausible priorities in case they discover one is better than our current list.

3. Building capacity: for similar reasons, it's useful to have experts in a wide range of issues to help prioritise further effort, and so we can act more quickly if priorities change.

4. Getting more people involved: by working in one area, you might meet people who want to get involved in other areas. This has happened a lot in effective altruism in the past.

The overall picture is that perhaps 50% should focus on the top issues, 30% on secondary issues, and 20% should be spread across perhaps 30+ other promising or important areas.

Here is some data on how our community is currently allocated, and what allocation we think might be ideal. We also have an article that discusses community coordination in detail.

If you see yourself as part of a community aiming to tackle social issues, whether the effective altruism or more broadly, then in order to choose a problem, ask yourself:

1. How would the community's efforts ideally be allocated over issues?
2. Where are the gaps, and how can I best move the allocation towards the ideal?
3. Where is my comparative advantage compared to other community members?

Likewise, the community as a whole can ask itself the same question: how would the world's efforts ideally be allocated, and how can we best move it towards that ideal?

So which issues do we think are most pressing?

Given the importance of this question, we dedicate a couple of articles to it in the key ideas series.

First, we introduce the case for working to reduce existential risk, and why now might be the most important century. Then we give our ranked list of issues.

We've also supported the development of the field of global priorities research, which tries to break down and answer the questions involved. For example, how should we allocate resources between reducing the risk of a disaster on the one hand, and preventing a more certain harm on the other? There's now an institute dedicated to this topic at Oxford.

Despite the vital importance of this research, there are still only dozens of researchers directly focused on it.

Ben Garfinkel was considering a physics PhD, but thought that since so many extremely smart people already do physics research, it would be hard to make a big contribution. He learned about global priorities research through a talk we gave, and after testing it out through a temporary position, decided to switch fields. He then developed important criticisms of arguments for prioritising AI.

This kind of criticism is exactly what we want to see more of. There's a good chance our views of which problems are most pressing are incomplete or mistaken in some ways. We want people to make the case for alternatives, and find issues we haven't even thought of yet.

Further reading on how to compare problems

First, you can learn more about the important, neglected, tractable framework:

- Here's a popular introduction.
- Here's a more technical discussion of a quantitative version of the framework.

As part of our career planning process, we have an article that guides you through a more comprehensive process for comparing problems.

Here is some more in-depth reading:

- Doing good together: coordinating as a community
- Ben Todd and Robert Wiblin discuss how much problems differ in effectiveness in our podcast on the key ideas series
- Crucial considerations and wise philanthropy by Nick Bostrom
- The moral value of information by Amanda Askell

The case for reducing existential risks

Click to read online

In 1939, Einstein wrote to Roosevelt:

> 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 risks[18] 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 risk. 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 has recently published *The Precipice: Existential Risk and the Future of Humanity* which gives an overview of the existential risks facing humanity today, and what we can do to reduce them.
>
> Sign up for our newsletter and we'll send you a free copy!

[18] 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.

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.[19]

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 one thousand 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.

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

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 risks 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. That's higher than most people's chances of being in a plane crash

[19] See footnote 3 here.

(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.[20]

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.[21]

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.[22]

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.[23]

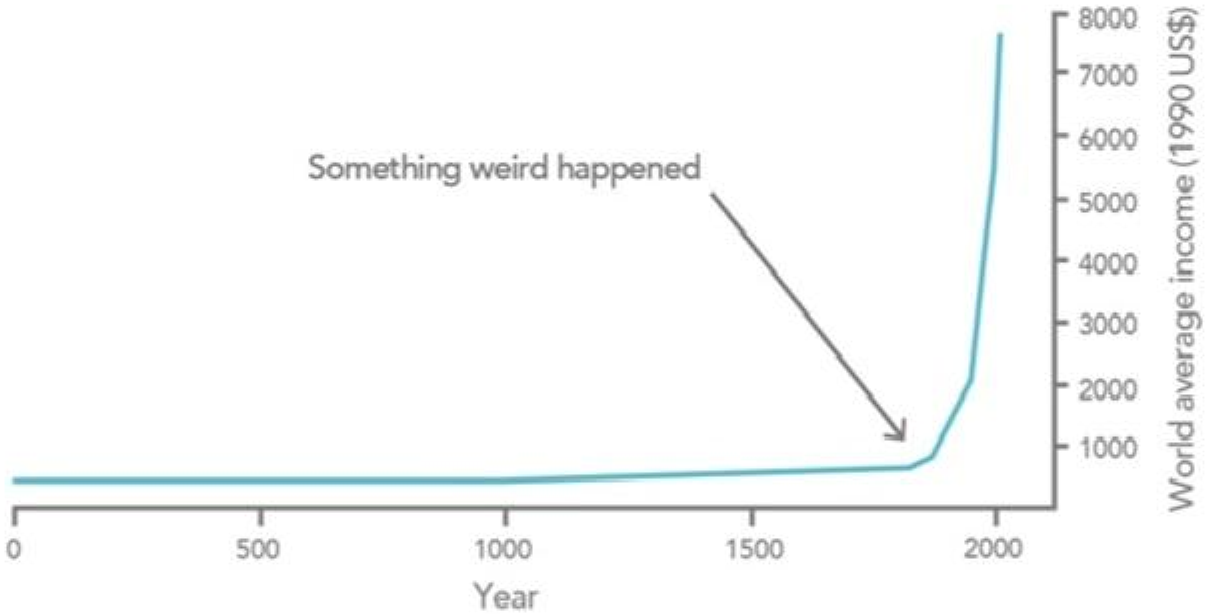
---

[20] The odds of crashing into the Atlantic on a Virgin operated A330 flying from Heathrow to JFK (1 in 5.4 million). So, you'd need to fly 1000 times in your life to be equally likely to be in a plane crash as in an asteroid disaster. "A crash course in probability," *The Economist*, 2015. Web link.

[21] A sufficiently large supervolcano could also cause a long winter that ends life. Some other natural risks could include an especially deadly pandemic, a nearby supernova or gamma-ray burst, or naturally caused runaway climate change.

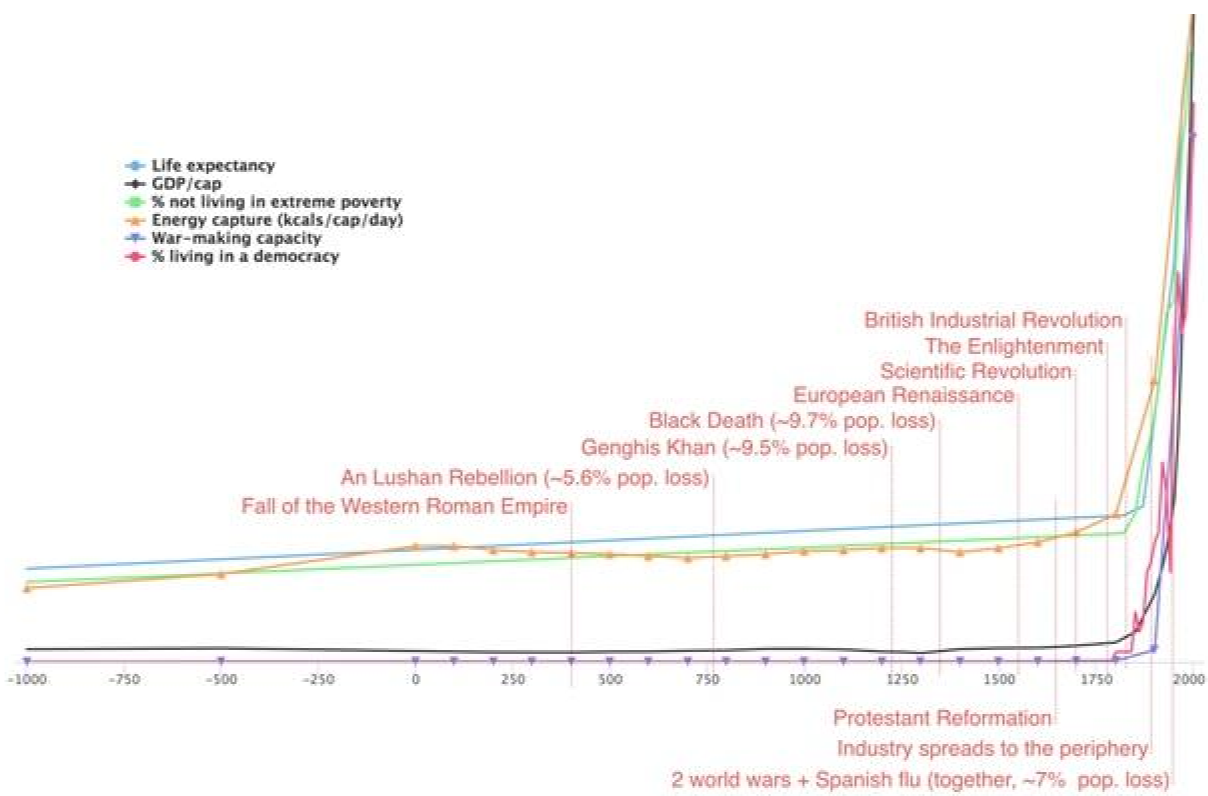
[22] You can see a summary of the contents of the paper in "Dr Toby Ord – Will We Cause Our Own Extinction? Natural versus Anthropogenic Extinction Risks", a lecture given at CSER in Cambridge in 2015. Link.

[23] Graph produced from Maddison, Angus (2007): *Contours of the World Economy, 1–2030 AD. Essays in Macro-Economic History*, Oxford University Press, ISBN 978-0-19-922721-1, p. 379, table A.4.



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.



Literacy and education levels have also dramatically increased:

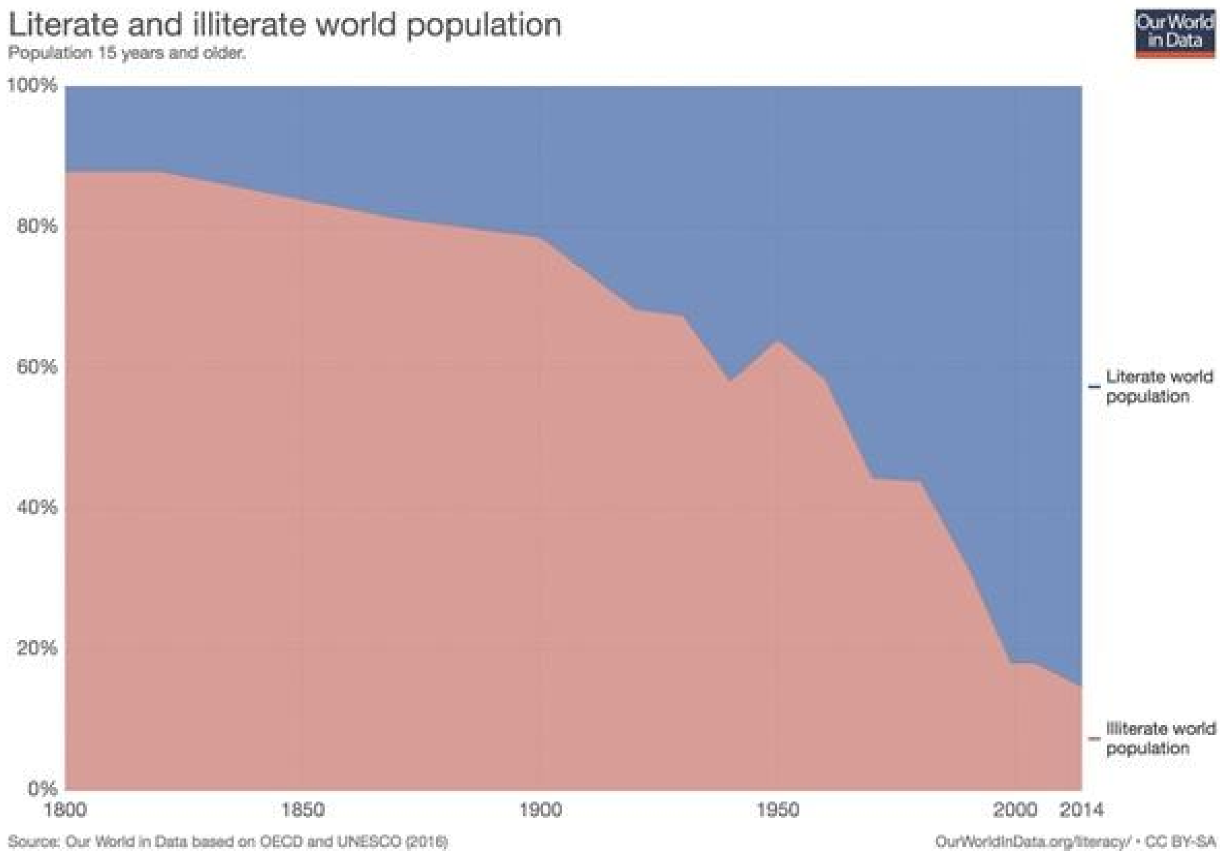


Image source

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.

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

Many people think the world is getting worse,[24] 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.

[24] Different surveys find substantially different results for how pessimistic people are about the future, but many find that a majority think the world is getting worse. For instance, a recent government survey in the UK found that 71% of respondents said they thought the world was getting worse. Pete Etchells, "Declinism: is the world actually getting worse?", *The Guardian*, 2015, Archived link, retrieved 17-October-2017.

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.[25] 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. 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.

---

[25] See footnote 12 here.

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."

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.[26]

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.

[26] For the chance of a bomb hitting a civilian target, see the figure one third down the page "What is the probability that a nuclear bomb will be dropped on a civilian target in the next decade?" Note that one expert estimated the chance of a nuclear strike on a civilian target in the next decade at less than 1%. *We're Edging Closer To Nuclear War*, Milo Beckman, FiveThirtyEight, 2017,
Archived link, retrieved 17-October-2017

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.[27]

Even a "mild" nuclear winter, however, could still cause mass starvation.[28] For this and other reasons, a nuclear war would be extremely destabilising, and it's unclear whether civilisation could recover.

How likely is a nuclear war to permanently end 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 of runaway 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."

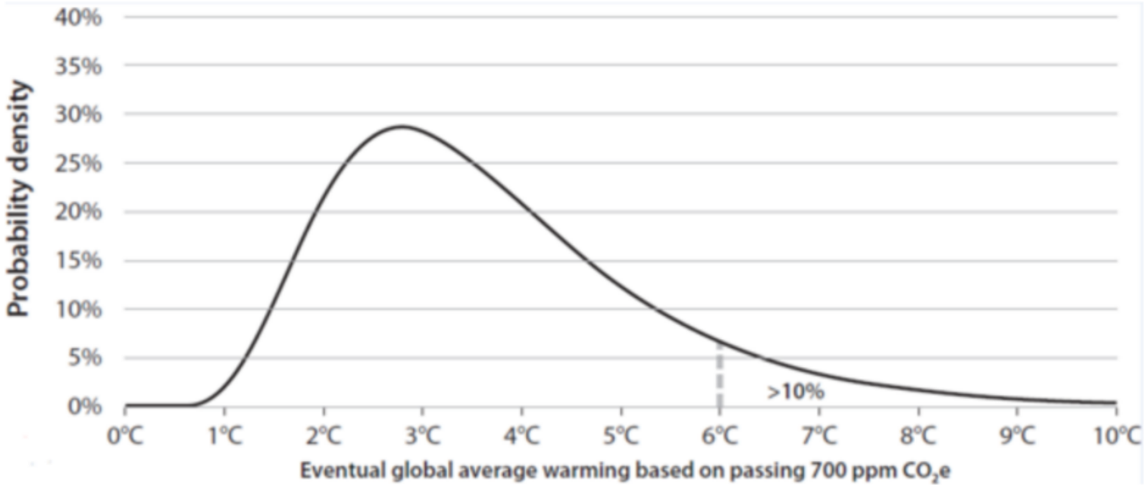
Climate change is certainly a major risk to civilisation.

The graph below shows estimates of climate sensitivity. Climate sensitivity is how much warming to expect in the long term if CO_2 concentrations double, which is roughly what's expected within the century.

---

[27] Climate models involve significant uncertainty, which means the risks could easily be higher than current models suggest. Moreover, the existence of model uncertainty in general makes it hard to give very low estimates of most risks, as is explained in: Ord, T., Hillerbrand, R., & Sandberg, A. (2010). "Probing the improbable: methodological challenges for risks with low probabilities and high stakes." *Journal of Risk Research, 13*(2), pp. 191-205. arXiv:0810.5515v1, Link.

[28] The expected severity of nuclear winter is still being debated, and Open Philanthropy recently funded further investigation of the topic.



An estimate of the likelihood of warming due to a doubling of greenhouse gas concentrations (Source: Wagner & Weitzman "Climate Shock" (via NPR)).

Image source

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

However, these estimates give a 10% chance of warming over 6 degrees, and perhaps a 1% chance of warming of 9 degrees. That would render large fractions of the Earth functionally uninhabitable, requiring at least a massive reorganisation of society. It would also probably increase conflict, and make us more vulnerable to other risks.

(If you're sceptical of climate models, then you should *increase* your uncertainty, which makes the situation *more* worrying.)

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

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, which should raise our estimate of the risks even higher. (Read more about runaway climate change.)

What new technologies might be as dangerous as nuclear weapons?

The invention of nuclear weapons led to the anti-nuclear movement just a decade 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.

In 1918–1919, over 3% of the world's population died of the Spanish Flu.[29] 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 biology PhDs. In fact, in 2006, *The Guardian* was able to receive segments of the extinct smallpox virus by mail order. Some terrorist groups have expressed interest in using indiscriminate weapons like these. (Read more about pandemic risks.)

Who ordered the smallpox? (Credit: The Guardian)

[29] See footnote 20 here.

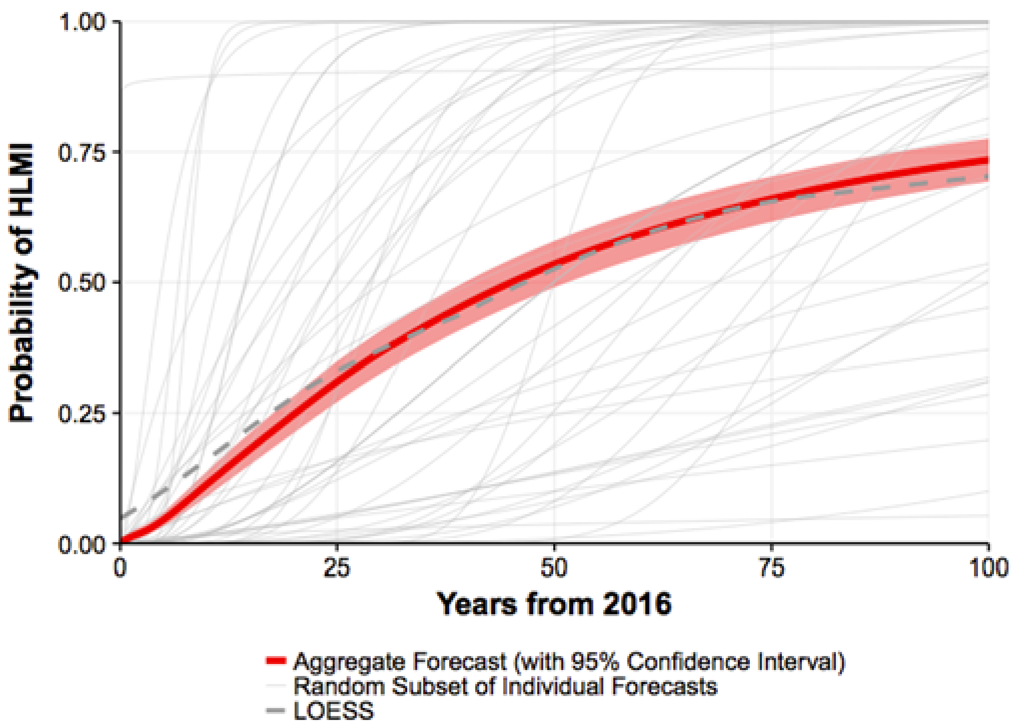
Another new technology with huge potential power is artificial intelligence.

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 have 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.



These probabilities are hard to estimate, and the researchers gave very different figures depending on precisely how you ask the question.[30] 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.

[30] For a discussion of the inconsistencies in the estimates, see the blog post released by AI Impacts, "Some Survey Results," Archived link, retrieved 30-Oct-2017.

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,[31] 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 mis-used. 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. 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.)

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

---

[31] See footnotes 15-18 in *Existential risk from artificial intelligence*, Wikipedia, archived link, retrieved 21-Oct-2018.

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%.

For instance, an informal poll in 2008 at a conference on catastrophic risks found they believe it's pretty likely we'll face a catastrophe that kills over a billion people, and estimate a 19% chance of extinction before 2100.

Risk	At least 1 billion dead	Human extinction
Number killed by molecular nanotech weapons.	10%	5%
Total killed by superintelligent AI.	5%	5%
Total killed in all wars (including civil wars).	30%	4%
Number killed in the single biggest engineered pandemic.	10%	2%
Total killed in all nuclear wars.	10%	1%
Number killed in the single biggest nanotech accident.	1%	0.5%
Number killed in the single biggest natural pandemic.	5%	0.05%
Total killed in all acts of nuclear terrorism.	1%	0.03%
Overall risk of extinction prior to 2100	n/a	19%

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

In our podcast episode with Will MacAskill, we discuss why he puts the risk of extinction this century at around 1%.

In his book *The Precipice: Existential Risk and the Future of Humanity*, Dr Toby Ord gives his guess at our total existential risk this century as 1 in 6 — a roll of the dice. (Listen to our episode with Toby.)

What should we make of these estimates? Presumably, the researchers 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). But does that mean we can dismiss their concerns entirely?

Given this, what's our personal best guess? It's very hard to say, but we find it hard to confidently ignore the risks. Overall, we guess the risk is likely over 3%.

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.[32]

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

---

[32] There are millions of deaths each year due to easily preventable diseases, such as malaria and diarrhoea, though the numbers are falling rapidly, so it seems unlikely they will exceed 300 million in the next century: Annual malaria deaths cut from 3.8 million to about 0.7 million, Annual diarrhoeal deaths (cut) from 4.6 million to 1.6 million. *Aid Works (On Average)*, Dr Toby Ord, Giving What We Can, Web.

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 civilization 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 ten 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 mind-blowing 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. 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:[33]

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.

[33] See footnote 28 here.

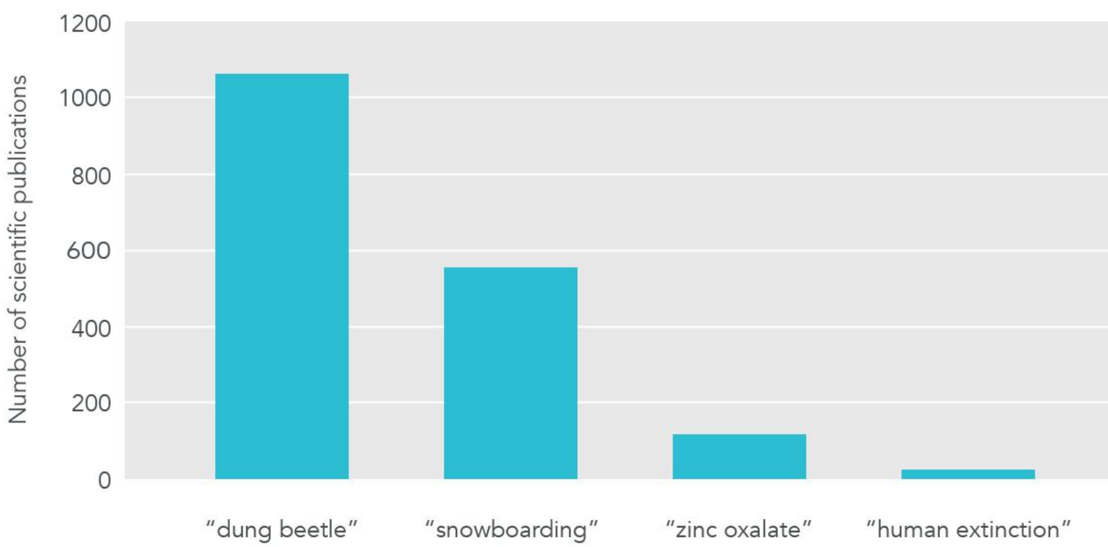
Far less is spent mitigating catastrophic risks from climate change. Welfare spending in the US alone dwarfs *global* spending on climate change.

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 $10 million 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. Runaway 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 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 to 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 contains 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 risk 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.

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). 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). So these estimates would put existential risk reduction as better or in the same ballpark 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 the Against Malaria Foundation.[34]

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 risks. 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 risks 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").

[34] If you donate $1 million over your life, about a third of the income of the mean college graduate, and we use the cost per life saved for Against Malaria Foundation, that would save about 130 lives. Source for income of college grads: Carnevale, Anthony P., Stephen J. Rose, and Ban Cheah. "The college payoff: Education, occupations, lifetime earnings." (2011). Archived link, retrieved 21-October-2017.

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." 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."[34]

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 risks, 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, then the value of reducing these risks goes down (or if we have much more obligation to reduce suffering than increase wellbeing). 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.

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.

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 combined the view that we don't have large obligations to future generations with the position 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, mental health or speeding up technology. Alternatively, you might think there's another moral issue that's more important, such as factory farming.

Want to help reduce existential risks?

Our generation can either help cause the end of everything, or be the generation that navigates 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 risks 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.

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.

Further reading

- Read about 51 policy and research ideas for reducing existential risk.
- See the academic version of this argument in this paper by Prof. 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 podcast with Carl Shulman on the common-sense case for existential risk work and its practical implications.

This could be the most important century

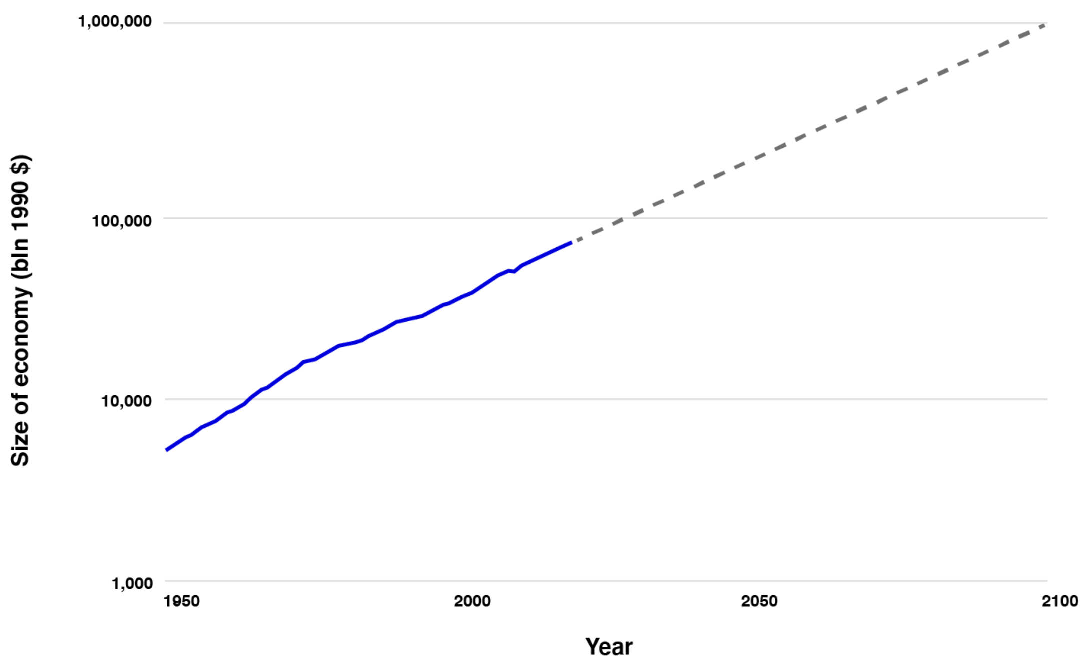
Click to read online

Will the future of humanity be wild, or boring? It's natural to think that if we're trying to be sober and measured, and predict what will really happen rather than spin an exciting story, it's more likely than not to be sort of… dull.

But there's also good reason to think that that is simply impossible. The idea that there's a boring future that's internally coherent is an illusion that comes from not inspecting those scenarios too closely.

At least that is what Holden Karnofsky — founder of charity evaluator GiveWell and foundation Open Philanthropy — argues in his new article series, "The Most Important Century."

The bind is this: for the first 99% of human history, the global economy (initially mostly food production) grew very slowly: under 0.1% a year. But since the Industrial Revolution around 1800, growth has exploded to over 2% a year.

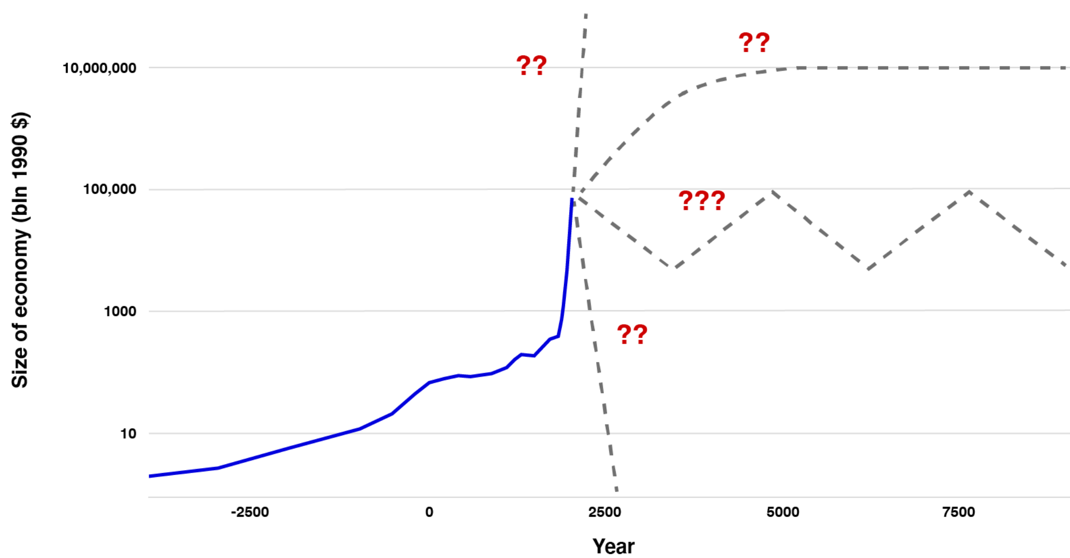


To us in 2020, that sounds perfectly sensible and the natural order of things. But Holden points out that in fact it's not only unprecedented, it also can't continue for long.

The power of compounding increases means that to sustain 2% growth for just 10,000 years — 5% as long as humanity has already existed — would require us to turn every individual atom in the galaxy into an economy as large as the Earth's today. Not super likely.

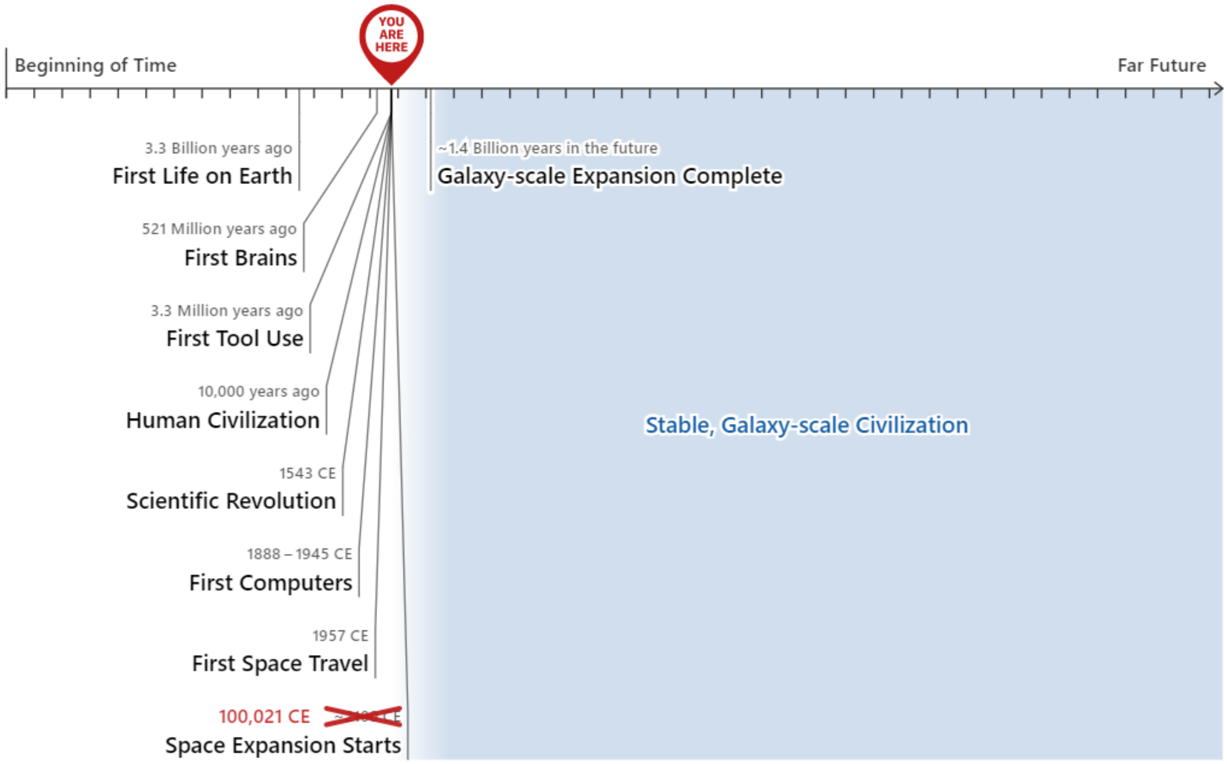
So what are the options? First, maybe growth will slow and then stop. In that case, we live today in the single miniscule slice in the history of life during which the world rapidly changed due to constant technological advances, before intelligent civilisation permanently stagnated or even collapsed. What a wild time to be alive!

Alternatively, maybe growth will continue for thousands of years. In that case, we are at the very beginning of what would necessarily have to become a stable galaxy-spanning civilisation, harnessing the energy of entire stars among other feats of engineering. We would then stand among the first tiny sliver of all the quadrillions of intelligent beings who ever exist. What a wild time to be alive!



Isn't there another option where the future feels less remarkable and our current moment not so special?

While the full version of the argument above has a number of caveats, the short answer is "not really." We might be in a computer simulation and our galactic potential all an illusion, though that's hardly any less weird. And maybe the most exciting events won't happen for generations yet. But on a cosmic scale, we'd still be living around the universe's most remarkable time:



Holden himself was very reluctant to buy into the idea that today's civilisation is in a strange and privileged position, but has ultimately concluded "all possible views about humanity's future are wild."

In the full series, Holden goes on to elaborate on technologies that might contribute to making this the most important era in history, including computer systems that automate research in science and technology, the ability to create 'digital people' on computers, or transformative artificial intelligence itself — and how they might create a world much weirder than most science fiction.

And if we simply project forward available computing power, or use expert forecasts, we can make a good case that the chance that these technologies arrive in our lifetimes is above 50%.

All of these technologies offer the potential for huge upsides and huge downsides. Holden is at pains to say we should neither rejoice nor despair at the circumstance we find ourselves in. His feeling is an "odd mix of intensity, urgency, confusion and hesitance." Going forward, these issues require sober forethought about how we want the future to play out, and how we might as a species be able to steer things in that direction.

If this sort of stuff sounds nuts to you, Holden gets it — he spent the first part of his career focused on straightforward ways of helping people in poor countries. Of course this sounds weird.

But he thinks that, if you keep pushing yourself to do even more good, it's reasonable to go from "I care about all people — even if they live on the other side of the world" to "I care about all people — even if they haven't been born yet" to "I care about all people — even if they're digital."

If this idea is correct, what might it imply in practical terms? We're not yet sure. You can see more of Holden's thoughts on the implications here in the series.

One consequence is that our actions might have huge stakes, making it even more important to reflect on where to focus.

Some specific priorities that seem helpful include:

- AI alignment research
- Global priorities research to give us more strategic clarity
- Building communities who take this idea seriously (e.g. the effective altruism community)
- Making it more likely that governments can make thoughtful, values-driven decisions

But most of all, we need to take this idea more seriously and understand its implications better.

To learn more, listen to our interview with Holden. You can also read or listen to the full series on Holden's blog:

- All possible views about humanity's future are wild
- The Duplicator
- Digital people people would be an even bigger deal
- This can't go on
- Forecasting transformative AI, part 1: what kind of AI?
- Forecasting transformative AI, part 2: what's the burden of proof?
- Are we "trending toward" transformative AI? (How would we know?)
- Forecasting transformative AI: the "biological anchors" method in a nutshell
- AI timelines: where the arguments, and the "experts," stand
- How to make the best of the most important century?
- Call to vigilance

Further reading

- Are we living at the most influential time in history? by Will MacAskill, with some reasons to be sceptical about the thesis. Also see our interview with Will about the same topic.
- Patient longtermism

- Forecasting TAI with Biological Anchors by Ajeya Cotra
- Podcast: Forecasting AI with Katja Grace
- New Report on How Much Computational Power it Takes to Match the Human Brain by Joseph Carlsmith
- Modeling the Human Trajectory by David Roodman
- Could Advanced AI Drive Explosive Economic Growth? by Tom Davidson
- Semi-Informative Priors Over AI Timelines by Tom Davidson

Our current list of pressing world problems

Click to read online

If you want to maximise your chance of having a big positive impact with your career, we think it's usually best to work on a global issue that's large in scale, solvable, and neglected. These are not always the biggest problems in the world — rather they are the issues that receive little attention *compared* to how important they are and how much can be done about them.

This page presents our current list of which world problems we think best fit this profile, and therefore seem most promising for more people to work on right now.

We begin this page with some categories of especially pressing world problems we've identified so far, which we then put in a roughly prioritised list.

We then give a longer list of global issues that also seem promising to work on (and which could well be more promising than some of our priority problems), but which we haven't investigated much yet.

Finally, we talk about what, in practice, these categories might mean for your career.

What is this page based on?

Our views draw on work by the University of Oxford's Global Priorities Institute, the Open Philanthropy Project, and our own research. Read about a framework we use for comparing issues, and the moral and methodological assumptions behind our views.

The most distinctive aspect of our approach is probably 'longtermism.' Longtermism is the idea that because such huge numbers of individuals might live in the long-run future, and because we think everyone's interests matter equally, approaches to improving the world should be evaluated mainly in terms of their potential for long-term impact — over thousands, millions, or even billions of years.

The kinds of issues we currently prioritise most highly

Emerging technologies and global catastrophic risks

In the 1950s, large scale production of nuclear weapons meant that a few world leaders gained, for the first time, the ability to kill hundreds of millions of people. This was a striking milestone in a robust trend: as technology improves and the world economy grows, it gets easier to cause destruction on an ever larger scale.

In the 21st century, we expect this trend to continue. New transformative technologies may promise a radically better future, but also pose catastrophic risks. Mitigating these risks, while increasing the chance these technologies allow future generations to flourish, may be the crucial challenge of this century.

There is a growing movement working to address these issues, including new research institutes at Cambridge, MIT, and Oxford. Nonetheless, work on mitigating many risks remains remarkably neglected — in some cases receiving attention from only a handful of researchers. If you can find an effective way to work on these issues, we think it may be the most valuable thing you can do.

Building capacity to explore and solve problems

Comparing global problems involves lots of uncertainty and difficult judgement calls, and there have been surprisingly few serious attempts to make such big picture comparisons. And there are many global issues we haven't yet seen investigated much at all.

For these reasons, we're also strongly in favour of work that might help resolve some of this uncertainty, as well as work that seems robustly useful on many different assumptions yielding different conclusions about what's best to work on.

One top priority in this category is to build the new field of 'global priorities' research, to try to work out which global problems are most pressing and make progress on foundational questions about how best to address them.

Another strategy is to help major existing institutions improve their capacities to make complex decisions, and therefore navigate global challenges.

A third strategy is to build communities of people who want to do good effectively, with the hope that they can deal with future challenges as they come. We're especially keen to build the effective altruism community, because it explicitly aims to work on whichever global challenges will be most pressing in the future. We count ourselves as part of this community because we share this aim.

Finally, we want to encourage people to explore problem areas where the case for impact is more speculative but which might be very pressing — e.g. promoting civilisational resilience, mitigating great power conflict, or laying the foundations for the governance of outer space. Making progress on these issues, which are less explored (especially from a longtermist perspective) can help establish and build these nascent fields, or help us discover they're less promising, meaning others can work more productively and efficiently.

Especially pressing global issues — our current overall list

If we had to rank the problem areas we've investigated so far in terms of the overall effectiveness of additional work on them (assuming someone had the same level of personal fit for each), our ranking would be as stated below.

Note that this is *not* a ranking of which world problems we think are the most important *full stop,* but rather a ranking of which problems we think are most pressing for people who broadly share our values and might follow our advice to work on *at the current margin*.

Partly for this reason, we expect this list to change to some extent year-to-year (perhaps seeing the addition or subtraction of one or two issues a year) as circumstances change — e.g. as some problems become less neglected or new issues arise.

Click through each of the links below to see our full writeups for each area — what the issue is and why we prioritise it as highly as we do.

Highest priority areas

- Positively shaping the development of artificial intelligence
- Global priorities research
- Building effective altruism
- Reducing global catastrophic biological risks

There are also many global issues we haven't looked into as much, but which seem like they could be as pressing as the issues above. We'd be excited to see a substantial minority of readers explore these issues in order to learn more about them, especially if they have unusually good fit for working on them. For instance:

- Mitigating great power conflict
- Global governance
- Governance of outer space
- See the rest of our list of other issues that could be top priorities

There are probably still other issues we haven't thought of, or perhaps an issue that stands above all the others we know of in terms of how pressing it is to work on — we call this 'cause X,' and it is another reason to do more exploration.

Second-highest priority areas

These are global problems we're confident are among the most pressing for more people to work on, but for which we'd guess an additional person working on them achieves somewhat less impact than work on our highest priorities, all else equal. Still, they could easily be someone's top choice depending on the circumstance.

- Nuclear security
- Improving institutional decision-making
- Climate change (extreme risks)

Together, all of the above categories make up what we call our **priority problem areas** or **priority problems**.

Other important global issues we've looked into

We'd love to see more people working on these issues, but given our general worldview they seem less pressing than our priority problems:

- Factory farming
- Health in poor countries

You can also see global issues we *haven't* looked into, but which also seem important to us (though likely less pressing than our priority problem areas) below.

Long lists of potentially pressing global issues beyond our current priorities

There are many global problems we have not yet looked into at length, but which, upon further investigation, might turn out to be very promising for people to work on. Below we list some issues we've at least briefly considered.

We'd be keen to see more of our readers gain expertise and test out projects in these areas than are currently doing so, especially within the first set of issues below. This is both because we think it might be directly valuable (especially for people with an unusually good personal fit for one of these areas or access to an unusually good opportunity), and because it presents a chance to discover new highly pressing issues.

For these reasons, we expect it makes sense for a significant minority of our readers (say 10–20%) to explore new areas like those listed below rather than focusing on our current priority problem areas. This would be the 10–20% who are relatively best suited to these areas, which probably means those with some kind of pre-existing interest. We discuss this more below.

Right now, we know very few people (who share our priorities) who are working on these areas, so if you find yourself choosing between one of the issues below and one of our highest-priority issues, and you have equally good opportunities and fit for each, we currently think working on one of these less explored but possibly very pressing issues could be your best bet.

We came up with this list by surveying seven advisors and combining their views with our own judgement. We've added some interesting sources with arguments about the importance of each area as leads to learn more, though we don't always agree with everything these sources say.

Note that the issues within each list are not in any particular order, and that there may be some overlap between them.

Potential highest priorities

The following are some issues that seem like they might be especially pressing from the perspective of improving the long-term future. We think these have a chance of being as important for people to work on as our priority problems listed above, but we haven't investigated them enough to know.

- Great power conflict
- Global governance
- Governance of outer space
- Voting reform
- Artificial sentience
- Improving individual reasoning or cognition
- Global public goods
- Surveillance
- Atomic scale manufacturing
- Broadly promoting positive values
- Civilisation resilience
- 'S-risks'
- Whole brain emulation
- Risks of stable totalitarianism
- Risks from malevolent actors
- Safeguarding liberal democracy
- Recommender systems at top tech firms
- We may need to invest more now to spend enough later on future problems

We'd be excited to see more discussion and exploration of many of these areas from the perspective of trying to improve the long-term future, and hope to help facilitate such exploration going forward.

Other longtermist issues

We're also interested in the following issues, but at this point think that work on them is likely somewhat less effective for substantially improving the long-term future than work on the issues listed above.

- Economic growth
- Science policy and infrastructure
- Harmful restrictions on migration
- Ageing
- Improving institutions to promote development
- Space settlement and terraforming
- Lie detection technology
- Wild animal welfare

Other global issues

We think the following issues are quite important from a short- or medium- term perspective, and that work on them might well be as impactful as additional work focused on reducing the suffering of animals from factory farming or improving global health.

However, we don't prioritise them as highly as those listed above because they seem somewhat less neglected, and because work on them seems less likely to substantially impact the very long-run future.

- Mental health
- Biomedical research and other basic science
- Increasing access to pain relief in developing countries
- Other risks from climate change
- Smoking in the developing world
- Other problems we've looked at which seem less impactful to work on than global health

Which issue should you focus on?

Determining for yourself which issues are most pressing

We gave our best guesses above about which global problems are most pressing for more people to work on. You might disagree, or want to do some of your own investigation before you make any decisions.

See this brief guide to investigating for yourself which problems are most pressing, which also links to a few other resources that can help you think things through.

Considering your personal fit

We encourage our readers to weigh how pressing a problem area is in general along with their personal fit for the area — how successful they are likely to be compared to the average person working on the problem, based on their skills and experience. You can think about your expected long-term impact in an area roughly as the product of how pressing the problem is that you're working on and how much you in particular will be able to contribute to solving it.

If you're coordinating as a part of a community, considering your comparative advantage — your fit for different areas compared to the community as a whole — may also be important.

For more on how to assess your personal fit and related topics, see our article Personal fit: why being good at your job is even more important than people think.

Advantages of spreading out over different issues

We do not think all our readers (let alone everyone) should work on our top ranked problems. Differing personal fit alone would mean that our readers should spread out over different problem areas, even if they were to all agree with our ranking of the general pressingness of the issues.

Moreover, as we gain more readers, there will be additional reasons for them to spread out. Two of the most important reasons are:

- As more people work on an issue, there are diminishing returns to additional work. This means that a group of people that's large compared to the capacity of an issue to absorb people will start to run out of fruitful opportunities to make progress on that issue, making it better for them to be spread out into other areas.
- If you work with others, there is value of information in exploring new world problems — if you explore an area and find out that it's promising, other people can enter it as well.

We cover this subject in more detail in our article on community coordination.

Among people who follow our advice, we aim to help a majority shoot for one of the highest-priority problem areas we listed above, but we'd also like to see a significant fraction aim for opportunities in the second and third groups. As we said above, we also think that some of our readers — perhaps 10–20% — will likely have the most impact by working on other global issues, especially those we list as potentially as pressing as our highest priority areas.

We think the reasons to spread out over different problem areas also apply to people aiming to take an 'effective altruism' approach to doing good — perhaps even moreso. For instance, if the effective altruism community becomes associated with a single issue, that could reduce its potential to grow and adapt in the future, which is an additional reason for people who take this approach to work on a variety of problem areas.

All that said, we think our highest-priority areas are currently neglected relative to how important they are, even within the effective altruism community, so we plan to continue to focus most of our efforts on them for now.

How do our organisational priorities overlap with these lists?

As a team, we also have limited capacity to provide advice and research. So, we try to focus a majority of our effort on our highest-priority areas. We then aim to put a smaller amount of effort into the second-highest-priority areas, and the remainder into other issues. For instance, most of our advising is focused on our highest-priority areas, but we also cover many other global issues on our podcast.

As our staff and readership grows, or if in the future our highest-priority problems become less neglected or we learn more about other pressing problems, we may prioritise a wider range of issues.

Other key articles on global priorities

- A framework for comparing global problems
- The story of how our views about which problem to focus on have changed over time
- How to compare global problems for yourself

4. Solutions: how can you find the most effective ways to tackle your chosen problem?

The best solutions are far more effective than others

Click to read online

Scared Straight was a government programme that received billions of dollars of funding, and was profiled in an award-winning documentary. The idea was to take kids who committed crimes, show them life in jail, and scare them into embracing the straight and narrow.

The only problem? One meta-analysis found the programme made the kids *more* likely to commit crimes, and another more recent meta-analysis found it had no effect.[35]

Causing this much harm is rare, but when social programmes are rigorously tested, a large fraction of them don't work.[36]

So, even if you've chosen a pressing issue, it would be easy to end up working on a solution to it that has very little impact.

Meanwhile, research finds that among solutions that do have positive effects, the best interventions within an area are *far* more cost effective than average, often achieving over 10 and sometimes 100 times as much for a given unit of resources.

In this article, we explain what we think the current research implies about how much solutions differ in effectiveness, why this should change how we approach making a difference, and how to find the best solutions within an area of practice.

---

[35] van der Put, Claudia E., et al. "Effects of awareness programs on juvenile delinquency: A three-level meta-analysis." *International Journal of Offender Therapy and Comparative Criminology*, vol. 65, no. 1, 1996, pp. 68-91. Archived link. Petrosino, Anthony, et al. "Scared Straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review." *Campbell Systematic Reviews*, vol. 9, no.1, (2013), pp. 1-55. Archived link.

[36] The percentage that work or don't work depends a lot on how you define it, but it's likely that a majority don't have statistically significant effects.

How much do solutions differ in how well they work?

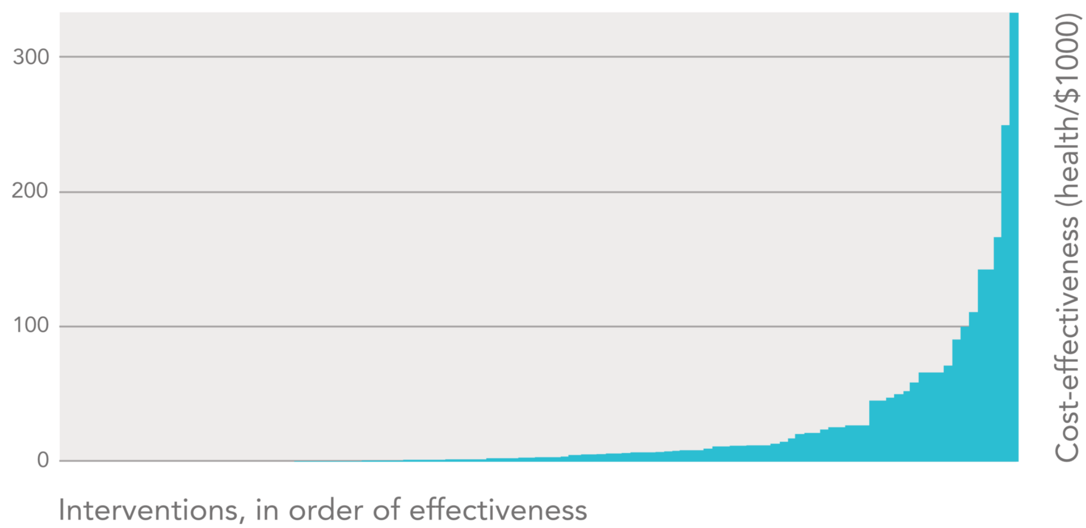
In recent years there's been a wave of advocacy to stop the use of plastic bags. However, convincing someone to entirely give up plastic bags *for the rest of their life* (about 10,000 bags) would avoid about 0.1 tonnes of CO2 emissions. In contrast, convincing someone to take just *one fewer* transatlantic flight would reduce CO2 emissions by over one tonne — more than 10 times as much.[37]

And rather than trying to change personal consumption in the first place, we'd argue you could do even more to reduce emissions by advocating for greater funding of neglected green technology.

This pattern doesn't just hold within climate change. Its significance was first pointed out in the field of global health by Toby Ord's article, The Moral Imperative toward Cost-Effectiveness in Global Health.

He found data that compared different health interventions in poor countries (e.g. malaria nets, vaccines, types of surgery) in terms of how many years of healthy life they produce per $1,000 invested.

This data showed that the most cost-effective interventions were around 50 times as cost effective as the median, 23 times the mean, and almost exactly obeyed the 80/20 rule.



Intervention cost effectiveness in global health in order of disability-adjusted life years (DALYs) per $1,000 on the y-axis, from the DCP2.

[37] According to the 2020 Founders Pledge Climate & Lifestyle Report, just one round trip transatlantic flight contributes 1.6 tonnes of CO2. Figure 2 of the same report shows the comparatively negligible effect of reusing plastic bags.

This is an incredible finding, because it suggests that one person working on the most effective interventions within global health could achieve as much as 50 people working on a typical intervention.

We've since seem similar patterns among:

- Large US social programmes
- Other global health datasets, such as WHO CHOICE and the DCP3.
- Public health in rich countries
- Education interventions in developing countries
- Ways to alter meat consumption to reduce animal suffering
- Policies to reduce CO2 emissions

In fact, it seems to show up wherever we have data.

There are some reasons to think that this data overstates the true differences in effectiveness between different solutions — especially those that you can actually work on going forward.

One reason is that often the very top solutions in an area are already being done by someone else.

A more subtle reason is regression to the mean. All the estimates involve a lot of uncertainty and error. Some interventions will 'get lucky' and end up with errors that make them look better than they are, and others will be unlucky and look worse.

In fact, even if all the solutions were equally effective, random errors would make some look above average, and others below average.

If we compare the interventions that appear best compared to the average, it's more likely that they benefited from positive errors. This means that if we investigate them more, they'll probably turn out worse than they seem.

This is what seems to have happened in practice. For instance, the data that Toby used found that deworming children was among the most cost-effective solutions in the dataset. However, the charity evaluator GiveWell discovered errors in the estimates in the study, and the studies behind them have been called into question in a debate that became known as the 'Worm Wars.'

But here's the final twist: in 2021, after 10 years of further research and scrutiny aiming to correct for these effects, GiveWell still recommends deworming charities as among the most cost effective in global health. This is true even though their best guess is that deworming is only about 10% as cost effective as the original estimates.

So, the original estimates were too optimistic, and overstated the spread from best to typical, but deworming still appears to be much more cost effective than average — likely still over 10 times better.

In fact, global health experts still believe that the best ways of saving lives in poor countries are around 100 times cheaper than the average.[38]

On the other hand, the data in these studies could also *understate* the true differences in effectiveness between solutions. One reason is that the data only covers solutions with easy-to-measure results that can be studied in trials, but the highest-impact ways of doing good in the past most often involved research or advocacy, rather than measurable interventions. This would mean the very best solutions are missing from the datasets.

For instance, a comparatively small number of people worked on the development of oral rehydration therapy, which now saves around one million lives per year. This research was likely extremely cost effective, but we couldn't directly measure its effectiveness before it was done.

Looking forward, we think there's a good case that medical research aimed at helping the global poor will ultimately be more cost effective than spending on direct treatments, increasing the overall degree of spread.

There's a lot more to say about how much solutions differ in effectiveness, and we'd like to see more research on it. However, our overall judgement is that it's often possible to find solutions that bring about 10 times as much progress per year of effort than other commonly supported solutions in an area, and it's sometimes possible to find solutions that achieve 100 times as much.

Technical aside: theoretical arguments about how much solutions differ

You might think that it's surprising that such large differences exist. But there are some theoretical arguments that it's what we should expect:

- There isn't much reason to expect the world of doing good to be 'efficient' in the same way that financial markets are, because there are only very weak feedback loops between having an impact and gaining more resources. The main reward people get from doing good is often praise and a sense of satisfaction, but these don't track the actions that are most effective. We don't expect it to be entirely inefficient either — even a small number of effectiveness-minded actors can take the best opportunities — but we shouldn't be surprised to find large differences.

[38] Caviola, Lucius, et al. "Donors vastly underestimate differences in charities' effectiveness." *Judgment and Decision Making*, vol. 15, no. 4, 2020, pp. 509-516. Link

- A relatively simple model can give a large spread. For instance, cost effectiveness is produced by the multiple of two factors; we'd expect it to have a log-normal distribution, which is heavy-tailed.

- Heavy-tailed distributions seem the norm in many similar cases, for instance how much different experts produce in a field, which means we shouldn't be surprised if they come up in the world of doing good.

- If we think there's some chance the distribution is heavy-tailed and some chance it's normally distributed, then in expectation it will be heavy-tailed, and we should act as if it is.

What do these findings imply?

If you're trying to tackle a problem, it's vital to look for the *very best* solutions in an area, rather than those that are just 'good.'

This contrasts with the common attitude that what matters is trying to 'make *a* difference,' or that 'every little bit helps.' If some solutions achieve 100 times more per year of effort, then it really matters that we try to find those that make *the most* difference to the problem.

This is why we highlight finding an effective solution as one of the four key drivers of your long-term impact, along with how pressing the problem is, how much leverage you have, and your degree of personal fit. It can be worth working on a less effective solution if that path does well on the other three factors, but it's one key thing to consider, especially once you reach the stage of your career where you're trying to directly tackle problems rather than build career capital.

So, how can we find the most effective solutions in an area?

Hits-based vs evidence-based approaches

There are two broad approaches among our readers:

1. **The evidence-based approach:** look for data about how much progress per dollar different solutions achieve, ideally randomised trials, and focus on the best ones.
2. **The hits-based approach:** look for rules of thumb that make a solution more likely to be among the very best in an area, while accepting that most of the solutions will be duds.

We generally favour the hits-based approach, especially for individuals rather than large institutions, and people who are able to stay motivated despite a high chance of failure.

Why? As noted, the best solutions typically can't be measured with trials, and so will be automatically excluded if you take the evidence-based approach. This is a serious problem because if the best solutions are far more effective than typical, it could be better to pick randomly among solutions that *might* be the very best, rather than to pick something that's very likely to be better than average but definitely not among the very best.

Another argument is that many institutions with social missions seem overly risk-averse. For instance, government grant agencies get criticised heavily for funding failures, but the employees at such agencies don't get much reward when they back winners. This suggests that individuals who are willing to take risks can get an edge by supporting solutions that have a high chance of not working. You can read more about the arguments for a hits-based approach.

One response to the hits-based approach is that it relies on deeply uncertain judgement calls, instead of objective evidence. That's true, but we contend that you can't escape relying on judgement calls. All our actions lead to ripple effects lasting long into the future. By taking an evidence-based approach, even if we suppose the evidence is fully reliable, at best you can measure some of the short-term effects, but you'll need to rely on judgement calls about the longer-term effects, which comprise the majority of all the effects. Learn more about cluelessness.

Given that judgement calls are unavoidable, the best we can do is to try to make the best judgement calls possible, using the best available techniques.

What does taking a hits-based approach involve in practice?

In short, we need to seek rules of thumb that make a solution more likely to be among the very best in an area (while unlikely to be negative). This could involve methods like the following:

- Upside/downside analysis: look for solutions with potentially very high upsides (and limited downsides).
- Apply the importance, neglectedness, and tractability framework, but at the level of solutions rather than problems. Here's an example of this kind of analysis within climate change by Founders Pledge.
- Make rough back-of-the-envelope estimates of progress per year. Even when the estimates are very uncertain, they can still help us spot very large differences

in effectiveness. You can further improve your estimates by applying best practices in forecasting. See some examples by the Open Philanthropy Project.
- Bottleneck analysis: try to identify the key step that would unlock progress in an area. For instance, in new areas, it's often most important to do direction-setting research and demonstrate progress is possible; then the priority becomes building a movement around the issue and scaling up the best solutions.
- Develop a theory of change.

In practice, this often ends up with a focus on research, movement building, policy change, or social advocacy, and on solutions that are unfairly neglected or seem unconventional.

In applying these frameworks, you can either try to do this analysis yourself, or find experts in the area who understand the need to prioritise and can do the analysis on your behalf (or a mix of both).

We're generalists rather than experts in the areas we recommend, so we mainly try to identify good experts — such as those on our podcast — and synthesise their views about how to tackle the problems we write about in our problem profiles.

If you aim your career at tackling a specific issue, however, then you'll probably end up knowing more about it than us, and so should put more weight on your own analysis.

Many of the areas we recommend are also small, so not much is known about how best to tackle them. This makes it easier than it seems to become an expert. It also means that your input on which solutions are best is especially valuable.

Following expert views also doesn't necessarily mean choosing 'consensus' picks, because those might fall into the trap of being pretty good but not best. Rather, if a minority of experts strongly supports an intervention (and the others don't think it's harmful), that might be enough to make it worth betting on. In brief, we aim to consider both the *strength* of views and *how good* the interventions seem, and are willing to bet on something contrarian if the upsides might be high enough.

For example, Sophie Rose switched to studying pandemic prevention due to our advice. When COVID-19 broke out, she realised human challenge trials could speed up vaccine development by many months, saving millions of lives. They also had a lot of public support, but weren't being permitted by regulators. So she co-founded 1DaySooner, which signed up volunteers for such trials, and one was eventually started in London in early 2021.

This didn't turn out to be fast enough to make a noticeable difference to the COVID-19 outbreak, but we think it was a bet worth taking.

More importantly, if there's another pandemic, 1DaySooner's work means human challenge trials could be ready to go right away, enabling us to develop vaccines far faster.

Further reading

- Is it fair to say most social programmes don't work?
- Hits-based giving — a blog post by Holden Karnofsky at Open Philanthropy.
- Improving global health in clear and direct ways — podcast with Alexander Berger from Open Philanthropy, where he compares evidence-based ways to reduce global poverty with hits-based ones. He claims that while the best hits-based approaches are sometimes better, they can't find enough of them to deploy their entire budget, so they also support evidence-based ones.
- Do social science findings generalise? — a podcast with Eva Vivalt
- Expected value estimates you can take (somewhat) literally — a blog post by Greg Lewis, which explains one way we might try to correct for regression to the mean.

Other key articles on effective solutions

- Is it fair to say that most social programmes don't work?

5. Leverage: which careers are highest impact?

What is leverage, and how can you get it?

Click to read online

People who want to have an impact often focus on jobs in which they help people directly — for instance, teaching and healthcare are two of the most common paths for college graduates.

So we worked with a medic, Greg Lewis, to estimate the number of lives saved by a typical clinical doctor. Greg estimated that the average doctor in a country like the US or UK adds several hundred years of healthy life over their career — equivalent to saving several lives. This is a lot of impact compared to most jobs, but it's less than many expect.

One reason is, as we've seen, issues like health in rich countries already receive a lot of attention. In this article, we'll discuss another reason: the impact of a clinical doctor is limited by the number of people they can treat with their own hands. In other words, this path has limited *leverage*.

By 'leverage,' we mean how many resources you're able to bring to bear on the best solutions to the most pressing problems.

Greg decided to switch to studying public health to research changes to government policy to prevent catastrophic pandemics. As a medic, he could respond to a pandemic by treating infected patients. But by working to improve pandemic policy, he can help make the efforts of thousands of doctors more efficient, and these efforts could also be directed to preventing another pandemic in the first place.

There are lots of career paths that give you more leverage, and we'll cover several more examples in this article.

The idea is not that the more indirect path is always better, but rather that you don't *have* to become a doctor, teacher, or charity worker to do good. By considering a broader range of ways to contribute, including indirect ones, you give yourself many more options, making it easier to find a path that's a good fit.

Also, more indirect paths can sometimes allow you to effect change at a greater scale, so it's *one* route to having a bigger impact.

Ways to increase your leverage

Once you've picked a problem to focus on and identified some promising solutions, you need to choose a career that will let you make them happen.

Some careers give you more leverage — and therefore let you put more resources toward great solutions than others.

To be a little more precise, when tackling a problem you can divide your impact into:

1. The effectiveness of the solution — how much progress you get per unit of resources invested.
2. Your leverage — how *many* resources you're able to get invested in the problem.

Your impact is given by the multiple of the two.

These resources could be your money, your own labour, the labour of others you enable, the budgets of large organisations you're able to influence, the power of a government you can help shape, and so on.

How much leverage you have in a job is usually a product of both the job itself and your personal fit.

The line between 'effective solutions' and 'leverage' is blurry. For instance, you can think of research either as a way of getting leverage or an especially effective type of solution. But it's normally useful to roughly divide them. If the best solution is research, you will want to consider whether you can get more research done by doing the research yourself or enabling other researchers to achieve more — e.g. as a research manager, by funding research, by lobbying the government to improve policy that influences research, and so on.

To illustrate this idea, it's perhaps easiest to give some examples of ways our readers have increased their leverage.

Improving government

Suzy Deuster wanted to become a public defender to ensure disadvantaged people have good legal defence. But she realised that while that path might improve criminal justice for perhaps hundreds of people over her career, by changing policy she might improve the justice system for thousands or even millions. She was able to use her legal background to start a career in policy, and now works in the Executive Office of the President of the US on criminal justice reform, and from there can explore other areas of policy in the future.

When we think of careers that 'do good,' we might not first think of becoming an unknown government bureaucrat. But senior government officials often oversee budgets of tens or even hundreds of millions — if you could enable those budgets to be spent just a couple of percent better, that would be worth millions of dollars of government revenue. And at the same time, government is often crucial in addressing many of the issues we most recommend people work on.

You can make a similar argument about helping to improve other large institutions, like philanthropic foundations or scientific grantmaking bodies. Most researchers want to focus on research, rather than administering grants, so these positions are often neglected. But grantmakers can influence how tens of millions of dollars are allocated — by improving that, you could enable more effective research to be done than you could achieve yourself.

Mobilising others

Suppose you've discovered an impactful job, but you're not sure you're a good fit. If you can instead find someone else to take it, then you've had just as much impact, if not more, compared to taking it yourself.

This is an example of being a *multiplier* — it's often possible to have a greater impact through enabling others than you can achieve directly. And that's another way to have leverage.

This was the original motivation for founding 80,000 Hours itself: we thought that if we could help just a couple of people have high-impact careers, that would do several times as much good as pursuing those careers ourselves.

While some jobs specialise in community building, you can pursue community building in any job — this illustrates why we shouldn't think of impactful jobs in terms of specific roles, but rather in terms of how you *best use* your role to do good. Learn more about how to do good in any job.

You can make a similar argument for careers in communication. By building an audience as a podcast host, journalist, or author, or by working in media, you can help spread important but neglected ideas to thousands of people — helping to mobilise their efforts or make them more effective.

Helping other people with leverage

If you know a person or an organisation that has leverage they're using to tackle a pressing problem, you can lend your skills to help further their impact.

For instance, Kyle had just graduated and was wondering about becoming a career counsellor or earning to give in the corporate world. Instead he decided to gamble on moving to Oxford to help the effective altruism research community there. He eventually became Nick Bostrom's assistant — a researcher he thinks is doing world-changing work. His thinking was that if he could save Bostrom 10% of his time on top of what he'd save with his next-best assistant, then he would enable him to perform 10% more research, contributing to the world-changing work.

Later Kyle moved into operations management, helping research organisations in the San Francisco Bay Area run more efficiently and grow faster. This also illustrates how helping to build organisations can be a route to leverage, since a well-run organisation can enable tens or hundreds of people to work together.

We list open positions in high-impact organisations on our job board, which need all kinds of skills, including marketing, research, data science, engineering, operations, personal assistants, management, and so on.

A special case of this kind is to help start new organisations. If you're able to build something that can hire people and continue without you, then you've multiplied your efforts — the organisation will be able to do more than you would ever be able to do with just your own hands.

Donating money

Another form of leverage comes from money: donations can be targeted at the most effective organisations in the world that are most in need of funding. You might not want to work at a nonprofit yourself, but by working in (for example) software engineering or accounting and donating some of your income, you might be able to fund the salaries of several nonprofit workers.

More broadly, if you have a very specific skillset — or don't want to change your career — by earning and donating money, you can 'convert' your skills into skilled labour working on the most pressing issues. The more you're able to donate, the more leverage this gives you.

We call this 'earning to give' — finding a career that uses your strengths and allows you to donate more, even if its direct impact is only neutral.

An extreme example of earning to give is Sam Bankman-Fried. He learned about the arguments for earning to give when he attended a talk by one of our founders while studying physics as an undergraduate at MIT.

Through others he met in our community, Sam found a job that used his mathematical skills in quantitative trading at Jane Street Capital, and that was a great fit. From there, he went on to help found cryptocurrency derivatives exchange FTX and build the decentralised finance movement.

Now, Forbes estimates Sam's net worth is $16 billion, making him likely the world's wealthiest person under 30. That's a lot of resources, making Sam's path so far very high leverage indeed. Sam has already donated millions to causes like animal welfare and the Biden campaign, and intends to donate most of his future wealth — enough to fund thousands of others doing high-impact work.

Sam took earning to give to the extreme. Going from a physics undergrad to a pioneer in decentralised finance in less than a decade, Sam is now worth billions — which he plans to give away.

Over 500 of our readers are also pursuing earning to give on a more modest scale. For example, John Yan decided that he could best contribute by staying in his current job (software engineering) and donating 10–30% of his income to effective charities. Collectively the contributions of these readers will add up to tens of millions of dollars in donations, which can do a huge amount of good.

Through research and technology

If you discover an important idea, it can be shared with everyone for free, meaning many more people can use their time better. The low marginal costs of spreading new ideas is one reason why helping to develop new ideas through research can be an impactful path.

Similarly, if you can develop a new technology that can be easily copied and shared, you can enable millions of others to achieve more.

You don't have to invent these ideas yourself to have a big impact — as we covered earlier, you can help more discoveries be made by building organisations, being a multiplier, and providing funding to support others in innovation.

Building your skills

You can also increase your leverage by developing skills that are more valuable to the issues you want to tackle.

One highly skilled person might cover the ground of several, giving them several times as much leverage.

You can increase the value of your skills by:

- Learning skills that are especially needed in the problem you want to work on
- Learning skills that fit you better — we cover personal fit elsewhere.
- Practising and training to increase your skills

Remember that getting leverage takes time. While many young people want to have a big impact right away, research finds that most people reach their peak output at age 40–60. We discuss how to increase your leverage by investing in your career capital in an upcoming article.

Conclusion

When aiming to do good, we typically first think of ways to contribute directly.

But often more indirect routes, like policy change or research, can let you reach a greater scale of impact.

And in fact, you can use almost any role to contribute to a pressing problem via community building, spreading important ideas, or donating. We normally think of 'social impact careers' in terms of specific job titles (e.g. working in corporate social responsibility or at a social impact organisation), but this shows that *how you use* your current role is at least as important as what role you have.

These more indirect paths don't always produce as much 'warm glow' as helping directly, but if you might be able to have a bigger impact, they're worth taking seriously.

These indirect paths might also fit you better in other ways, like being more intellectually interesting, or letting you work with colleagues you like. The warm glow of helping directly is not the only way to make a career satisfying and meaningful.

In general, by considering a much wider range of ways to contribute, it's often possible to find a path that's better overall — both for you and in terms of impact.

Our list of high-impact careers

Click to read online

The highest-impact career for you is the one that allows you to make the biggest contribution to solving one of the world's most pressing problems. On this page, we list some broad categories of impactful careers, followed by about 30 more specific and unusual career paths we think are especially impactful, such as long-term AI policy research. The lists are based on 10 years of research and experience advising people, and represent the careers it seems to us will be most impactful over the long run if you get started on them now — though of course we can't be sure what the future holds.

You can use the lists on this page to get new ideas for impactful careers and make sure you haven't missed a great option. Then select between them primarily based on your fit — see our guide on how to make a career plan for details on how to choose. Click on the profiles to learn more about why we chose each one, how to assess your fit, and see open high-impact job opportunities.

Key categories of impactful careers

Before thinking about specific career paths, we think it's valuable to consider what *kinds* of careers tend to be highest impact. The career categories below can enable you to make a big contribution to whichever global problems you think are most pressing.

Government and policy in an area relevant to a top problem

Work out how government policy can help solve the world's most pressing problems, and help make those policies happen. Read more

Organisation-building at effective nonprofits

Help build great organisations doing important work via entrepreneurship, operations, people management, project management, fundraising, or administration.
Read more

Research in relevant areas

Aim to make intellectual advances about how to tackle the world's most pressing problems. Read more

Applying an unusual skill to a needed niche

We mostly focus on more common skills, but a huge variety is needed. If you already have expertise in a narrow area, there might be a way to apply it to a pressing global problem. Read more

Communication

Convey important ideas and information in a compelling way, and you can help others focus on the right things and work more effectively. Read more

Earning to give

Take a job that fits you well and lets you contribute financially to funding-constrained, highly effective organisations. Read more

List of top-recommended career paths

If you want to help tackle the global problems we think are most pressing, these are the career paths we most recommend — our **priority paths**. Most of them are difficult to enter — you may need to start by investing in building skills for several years, and there may be relatively few positions available. However, if you have the potential to excel in any of these paths, we encourage you to seriously consider it.

We've ranked these paths roughly in terms of impact, assuming your personal fit for each is constant. But there is *a lot* of variation within each path — so the best opportunities in one lower on the list will often be better than most of the opportunities in a higher-ranked one.

1. AI safety technical research
   The development of AI could transform society. Help increase the chance it's positive by tackling the technical problems of AI alignment, such as 'corrigibility,' 'interpretability,' and reliability.
2. Long-term AI policy strategy research and implementation
   The deployment of powerful AI systems could go very well or very badly for society. Help shape and implement policies to make it go well.
3. Founder of new projects tackling top problems
   Most people can only do great work within a great organisation. Make it possible to deploy more talent and resources toward progress on pressing global problems.
4. Grantmaker focused on pressing world problems
   Allocate philanthropic funding as effectively as possible by identifying and vetting new organisations and projects.
5. Helping build the effective altruism community
   If you could cause two other, equally talented people to use their skills to tackle the world's most pressing problems, that'd be double the impact you'd have if you did it yourself.
6. Operations management in high-impact organisations
   One of the biggest bottlenecks for organisations working on pressing global problems is excellent operations staff to design, scale, and implement great systems.

7. Research into global priorities
   Help identify the most pressing global problems and the most effective ways to solve them to enable others to have a greater impact.
8. Biorisk research, strategy, and policy
   Help reduce the risk of a global biological catastrophe, like an engineered pandemic much worse than COVID-19.
9. China-related AI safety and governance paths
   Help Chinese companies and stakeholders involved in building AI make the technology safe and good for society.
10. Forecasting and related research and implementation
    Help powerful institutions make good predictions and decisions, particularly around catastrophic risks.

Sometimes recommended: other high-impact career paths we're excited about

Below we list some other career paths that we don't recommend as often or as highly as those above, but which can still often be top options for people we advise. Take a look and consider any that might be a good fit for you. These aren't ranked in terms of impact.

High-impact but especially competitive

- Historian of large societal trends, inflection points, progress, or collapse. Help anticipate future changes to society and technology by examining the past.
- Public intellectual. Build a platform and spread important ideas about pressing global problems and how we can best solve them.

Potentially high-impact but still under-researched

- Expert in AI hardware. Shape the course and pace of AI deployment though AI hardware — e.g. by restricting or allowing access to compute by different actors.
- Information security. Help prevent emerging technologies like AI and biotech from being misused, stolen, or tampered with.
- Policy careers focused on other pressing global issues. Research, promote, or implement policies that address global issues beyond biosecurity and AI safety.
- Specialist in emerging global powers. Help emerging global powers coordinate with the rest of the world in addressing pressing global problems.
- Manager of a long-term philanthropic fund. Help philanthropists invest resources now in order to deploy the returns in the medium- to long-term future, or whenever the time is right.

- Investigate a potentially pressing but unexplored global issue. See if we can make progress on an under-researched problem that could be even more pressing than those we currently prioritise.

Other impactful options if you're an especially good fit

- Journalism. Spread important ideas and help shape public discourse for the better.
- Non-technical roles in leading AI labs. You don't have to be a technical or policy researcher to help AI labs work toward greater safety.
- Software engineering. Use coding skills to help shape the development of AI, prevent pandemics, and support the most impactful nonprofits.
- Research management. Multiply the impact of research by guiding, coordinating, and promoting the best work.
- Organise an effective altruism local group. Grow the effective altruism community by running events, mentoring, and connecting people to one another.
- Think tank research. Improve government policies through targeted research and advocacy.
- Found a tech startup. Get great career capital by working at a tech startup — and if it works out, have a positive impact by donating part of your earnings.
- Academic research. Do research in academia, which is home to some of the most far-reaching and cutting-edge research we can do.
- Mitigate climate change using effective altruist approaches. Identify and implement the most effective solutions to climate change by focusing on the most extreme risks and neglected technologies.
- Earn to give in a high-paying role, such as quantitative trading. Get an extremely high-paying job, then donate some of your salary to the most effective charities you can find.
- Become an executive assistant for someone doing especially high-impact work. Know of someone doing very valuable work? Multiply their impact by taking time-consuming tasks off their plate.
- Improve China-Western coordination on global catastrophic risks. The broader version of our top-recommended path on shaping China's involvement in AI: help coordinate China and the West on a variety of global issues.
- Engineering. Develop and deploy potential technological solutions to important problems.

Steps to build career capital and increase your future impact

These aren't career paths per se, but they can be great phases in your career that can set you up to have an even greater impact later. This list is especially incomplete compared to the others.

- Congressional staffer. Learn about policy making and government while having a direct social impact by helping improve government policies.
- Economics PhDs. Get a valuable credential in a field relevant to most of our priority paths (especially global priorities research), plus set yourself up for good earning-to-give options.
- Machine Learning PhDs. Understand the cutting edge of machine learning in order to contribute to the safe deployment of AI.
- Early-stage startup employee. Learn about founding a startup while gaining connections and a broad skillset.

Other career reviews

Below are all the other career reviews we have written so far, listed alphabetically. Some of this content is a little out of date, so take the advice with a pinch of salt.

- Actuarial science
- Allied health professional
- Biomedical research
- Computer Science PhD
- Data science (for skill-building & earning to give)
- Executive search
- Founding effective nonprofits (international development)
- Front office finance (for skill-building & earning to give)
- Management consulting (for skill-building & earning to give)
- Marketing (for skill-building & earning to give)
- Medical careers
- Nursing
- Party politics in the UK
- Philosophy academia
- Product manager in tech
- Programme manager in international organisations
- Pursuing fame in art and entertainment
- Teaching
- UK commercial law (for earning to give)
- Web designer
- Working at effective nonprofits (international development)

How to choose which career path *you* should pursue

The purpose of these lists is to give you more ideas about high-impact career paths. There are likely many other options worth considering for your personal list that we don't cover.

As we cover in our key ideas series, the overall impact of your career depends on both:

- How impactful the path is in general
- Your degree of personal fit with it

So once you have some good ideas, we recommend you narrow down based on fit — including whether the path is likely to be sustainable and personally satisfying.

To help you work out which career path is best for you, we've created an eight-week planning course to help you systematically think through important factors in choosing a career, generate more ideas, and identify your best option. (After that, check out our job board for opportunities in the career path that fits you best.)

Other key articles on leverage

- How to have a significant impact without changing your job

6. Personal fit: what are you good at?

Personal fit: why being good at your job is even more important than people think

Click to read online

"Find work you're good at" is a truism, but we think many people still don't take it seriously enough.

Finding the option where you have the best chance of excelling over the course of your career — where you have your greatest 'personal fit' — is one of the key determinants of your career's impact. In fact, after initially identifying some promising paths, we think it's often the *most* important factor.

The first reason is that in many fields, data suggests that success is distributed unevenly.

This is most pronounced in complex jobs like research or entrepreneurship. A key study of 'expert performance' concluded:

> *A small percentage of the workers in any given domain is responsible for the bulk of the work. Generally, the top 10% of the most prolific elite can be credited with around 50% of all contributions, whereas the bottom 50% of the least productive workers can claim only 15% of the total work, and the most productive contributor is usually about 100 times more prolific than the least.*

In the most skewed fields like these, your expected impact is roughly just the value of outsized success multiplied by its probability — from an impact point of view, you can roughly ignore the middling scenarios.

But in most jobs there are still sizable differences in output between, say, the top 20% of performers and the average performers.

It's unclear how predictable these differences are ahead of time, and people often overstate them. But even if they're only a little bit predictable, it could matter a great deal — having slightly higher chances of success could result in large increases in impact.

For instance, suppose in option A you expect to be average, and in option B you expect to be in the top 30%. If the top 30% produce two times as much as average, then it could be better to take option B, even if you think option A is up to two times higher-impact on average.

If we also consider you'll be less replaceable if you're in the top 30%, the difference in counterfactual impact could be even larger.

The second reason why personal fit is so important is that being successful in almost any field gives you more connections, credibility, and money to direct towards pressing problems — increasing your career capital and leverage.

If you succeed at something, that gives you a reputation and credentials you can use to find future opportunities. You'll also tend to meet other successful people, improving your connections. And you might gain a platform or money you can use to promote neglected issues. This idea is discussed more in our podcast with Holden Karnofsky.

Being good at your job is also one of the main ingredients of a satisfying job, which helps you stay motivated in addition to being important in itself. It could easily be more satisfying to be in the top 20% of a profession, even if it's perhaps lower paid or less glamorous than an alternative where you'd be average.

How important is personal fit compared to other factors?

You can think of your degree of personal fit with a career option as a multiplier on how promising that option is in general, such that:[39]

> total impact = (average impact of option) x (personal fit)
>
> *and*
>
> total career capital = (average career capital) x (personal fit)

This means that we often advise people to first identify some high-impact paths, and then choose between them based on their degree of fit with them — especially focusing on those where they might excel.

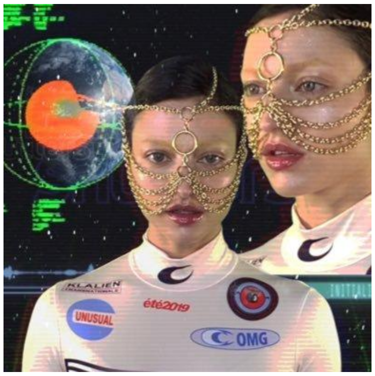
It also means that it can be worth taking a job that you think is, say, in your second tier for impact, but is a better fit for you.

Because personal fit is so important, we would almost never encourage you to pursue a career you dislike. Succeeding in almost any career takes many years and sometimes decades of work. If you don't like your job, you're unlikely to stick with it that long, and

[39] More specifically, we define a person's 'personal fit' for a job as the ratio between: 1) the productivity that person would have in the job in the long term, and 2) the average productivity of other people who are likely to take the job.

so you'll forgo a lot of your impact. (And there are other reasons we wouldn't encourage you to pursue a career you dislike.)

Although it's not what we most commonly recommend, it can sometimes even be worth taking jobs that don't have any direct connection to a particularly impactful path in the short term because of the career capital you might get from excelling in them.



Isabelle Boemeke started out as a fashion model, but after speaking to experts who said nuclear energy was needed to tackle climate change (but were afraid to promote it due to its unpopularity), she pivoted to using her social media following to promote it. Becoming a fashion model isn't normally one of our recommendations, but it could still be the right choice if your fit is high enough.

More generally, since you can have a significant impact in any job by donating, through political advocacy, or being a multiplier on others, simply working hard and being more successful in any path can let you have more impact.

What am I good at?

Academic studies and common sense both suggest that while it's possible to predict people's performance in a path to some degree, it's a difficult endeavour.[40] What's more, there's not much reason to trust intuitive assessments, or career tests either.[41]

So what does work?

Making predictions

Here are some questions you can use to make some initial assessments of your fit from several different angles:

[40] The best study we've found showed that the best predictors of job performance only correlate about 0.5–0.65 with job performance. This means that much of the variance is unexplained, so that even a selection process using the best available predictors will appear to regularly make mistakes. This matches personal experience; it's pretty common for hiring processes to make the wrong call and for new hires to not work out. Schmidt, Frank L., et al. "The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years…" Fox School of Business Research Paper, 2016, 1-74. PDF

[41] Most career tests are based on 'interest-matching,' often using a system similar to Holland types. However, meta-analyses have found that these methods don't correlate or only very weakly correlate with job performance. We cover some studies about this here.

1. **What do you think are your chances of success?**[42] To do this, look at your track record in similar work and try to project it forward. For instance, if you were among the top 25% of your class in graduate school, because roughly the top half of the class continue to academia, you could roughly forecast being in the top 50% of academia.[43] To get a better sense of your long-term potential, look at your rate of improvement rather than recent performance. (More technically, you can try to make a base rate forecast).
2. **What drives performance in the field, and how do you stack up?** The first step gives you a starting point, but you can try to improve your estimates by asking yourself what most drives success in the field, and whether you have those traits, as well as looking for other predictors of performance.
3. **What do experts say?** If you can, ask people experienced in the field for their assessment of your prospects. Just be careful not to put too much weight on a single person's view, and aim to ask people who have had experience selecting people for that job in question, and are likely to be honest with you.
4. **Does it match your strengths?** One way to gauge this is to look for activities that don't feel like work to you, but do for most people. We have an article about how to assess your strengths.
5. **Do you feel excited to pursue it?** Gut-level motivation isn't a reliable predictor of success, but if you *don't* feel motivated, it'll be challenging to exert yourself at the level required for high performance in most jobs. So a lack of excitement should give you pause.
6. **Will you enjoy it?** To stick with it for the long term, the path would ideally be reasonably enjoyable and fit with the rest of your life (e.g. if you want a family, you may want a job without extreme working hours).

Learning to make good predictions is an art, and one that's very useful if your aim is to do good, so we have an article about how to get better at it.

Investigating your options

Many people try to figure out their career from the armchair, but it's often more useful to go and test things in the real world.

[42] If the outcome of a choice of career path is dominated by 'tail' scenarios (unusually good or bad outcomes), which we think it often is, then you can approximate the expected impact of a path by looking at the probability of the tail scenarios happening and how good/bad they are.
[43] If we suppose that the 50% with the best fit continue to academia, then you'd be in the top half. In reality, your prospects would be a little worse than this, since some of your past performance might be due to luck or other factors that don't project forward. Likewise, past failures might also have been due to luck or other factors that don't project forward, so your prospects are a bit better than they'd naively suggest. In other words, past performance doesn't perfectly predict future performance.

If you have time, the next stage is to identify key uncertainties about your fit, and then investigate those uncertainties.

It's often possible to find low-cost ways to test out different paths. Start with the lowest-cost ways to gain information first, creating a 'ladder' of tests. For example, one such ladder might look like this:

1. First read our relevant career reviews and do some Google searches to learn the basics (1–2h).
2. Then speak to someone in the area (2h).
3. Then speak to three more people who work in the area and read one or two books (20h). You could also consider speaking to a career advisor who specialises in this area.
4. Then look for a project that might take 1–4 weeks of work, like applying to jobs, volunteering in a related role, or starting a blog on the policy area you want to focus on. If you've done the previous step, you'll know what's best.
5. Only then consider taking on a 2–24 month commitment, like a work placement, internship, or graduate study. At this point, being offered a trial position with an organisation for a couple of months can also be an advantage, because it means both parties will make an effort to quickly assess your fit.

In our planning process, we lead you through the process of identifying key uncertainties for each stage of your career, and then making a plan to investigate them.

If at any point you learn that a path is definitely not for you, then you can end the investigation.

Otherwise, when your best guess about which path is best stops changing, then it's time to stop doing tests and take a job for a few years. But that is also an experiment, just on a longer time scale — as we discuss in our article on exploration.

Further reading

- Holden Karnofsky on building aptitudes and kicking ass
- How much do people differ in productivity? What the evidence says.
- How to improve your judgement
- How to judge your chances of success
- How to identify your personal strengths: A collection of the best advice and research I've seen
- Don't go with your gut
- What makes for a dream job?
- Balancing impact and happiness
- How to analyse replaceability

- *Peak: Secrets from the New Science of Expertise*, by Anders Ericsson, makes the argument that success is mainly driven by years of focused practice. We think his conclusions are too extreme, but it's a provoking book, and the central idea — that attaining high levels of performance requires a lot of practice and it's possible to improve most of our skills — seems correct. Also see this nice summary of Ericsson's career by Cal Newport.

Other key articles on personal fit

- How to identify your personal strengths: collection of the best advice and research we've seen

7. Strategy: how to find your best career

Career capital: how best to invest in yourself

Click to read online

A key strategic consideration is 'career capital' — the skills, connections, credentials, and financial resources that can help you have a bigger impact in the future.

Career capital is potentially a vital consideration because people seem to become dramatically more productive over their career. Our impression is that most people have little impact in their first couple of jobs, while productivity in most fields seems to peak at age 40–50.[44] This suggests that by building the right career capital, you can greatly increase your impact, and that career capital should likely be one of your top considerations early in your career.

This leaves the difficult question of which options help you gain the best career capital, putting you in the best position to take the highest-impact roles addressing the world's most pressing problems.

How to gain the best career capital?

We've noticed that people often think the best way to gain career capital is by doing prestigious jobs, such as consulting. We think consulting is a good option for career capital, but it's rarely the most direct route into our priority paths.

You can see our write ups of individual priority paths for our thoughts on the best next steps to gain career capital within those paths. Some options stand out as good for a variety of paths and also have reasonable backup options:

- Go to **graduate school in a subject that provides a good balance of personal fit, relevance, and backup options**, with the aim of working in policy or doing relevant research. We'd especially highlight graduate study in economics and machine learning, since they provide good career capital and have great backup options, but some other useful subjects include: security studies, international relations, public policy, relevant subfields in biology, and more. Start with a master's, and only add a PhD if you plan to focus on research.
- Work as a **research assistant at a top think tank**, aiming to specialise in a relevant area of policy, such as technology policy or security — especially if you'd be working under a good mentor.

[44] Income usually peaks in the 40s, suggesting that it takes around 20 years for most people to reach their peak productivity. See the footnote on our Career capital page for more discussion.

- Take other **entry routes into policy careers**, such as (in the US) certain congressional staffer positions, joining a congressional campaign, or working directly in certain executive branch positions. In the UK, the equivalent would be working in the civil service, or working for a politician.
- **Work at a top AI lab**, including in certain non-technical roles. This potentially gives you similar career capital benefits to consulting, but with more relevance to AI. Certain jobs in 'big tech' can also be more attractive than consulting, especially if you can work in a relevant area or develop a useful skillset, such as AI or information security.
- Join a **particularly promising startup**, especially in the for-profit tech sector, though small and rapidly growing organisations in any sector are worth considering. Look for an organisation with impressive people, and a role that lets you develop concrete and relevant skills (especially management, operations, entrepreneurial, general productivity, and generalist research skills). It's also ideal to join an organisation working within a relevant area, such as AI or bioengineering. These roles may help you build relevant skills and connections, and also give you the opportunity to advance quickly. If you can build a network of people working in startups, then you can also try to identify an organisation that's unusually credible (e.g. backed by a range of impressive funders), and which might be on a breakout trajectory, which would give you the chance of further upside (e.g. reputation, money).
- **Do something that lets you learn about China**, such as taking the steps listed here.
- **Work at a top nonprofit or research institute in a top problem area**, such as some of those we list in our recommended organisations.
- Take any option where you might be able to have **unusually impressive achievements**. For instance, we came across someone who had a significant chance of landing a national TV show in India as a magician and was deciding between that and… consulting. It seemed to us that the magician path was more exciting, since the skills and connections within media would be more unusual and valuable for work on pressing problems than those of another consultant.
- If you might be able to do **something with significant positive impact in the next five years** (such as founding a new nonprofit), that can often be a great choice — not only is it impressive, but it also gives you connections and skills that are highly relevant to solving the problem you're working on.

Specialist vs transferable career capital

One common theme in the above is that, surprisingly often, there is little tradeoff between getting career capital and doing the natural first step towards top long-term roles.

However, there can be some tradeoff between gaining 'specialist' career capital versus 'transferable' career capital.

- Transferable career capital is relevant in lots of different options. For example, management skills (which are needed by almost every organisation), or achievements that are widely recognised as impressive.
- Specialist career capital, like knowledge of and connections within a specific global problem, prepares you for a narrow range of paths, but is often necessary to enter the highest-impact options.

Our overall view is that the rewards of specialisation are often high enough to make the costs worth it, especially if you also maintain some backup options while specialising.

Specialising increases the risk that if the situation changes, your skills could become less useful. If you're early in your career and/or very uncertain about which long-term options are the best fit for you, it can be better to focus on transferable career capital and plan on specialising later. Indeed, there's a chance the best option for you is something you haven't even thought of yet, and gaining transferable career capital is the best way to prepare for that.

Should you wait to have an impact?

Later in your career, you're more likely to have the option to take a job with immediate impact right away. At this point, how to value gaining career capital versus having an immediate impact becomes a much harder issue to settle.

For instance, if you could work as an AI safety engineer today, should you still do a PhD to try to open up potential research positions you think might be higher impact?

If you do the PhD, not only do you give up the impact you would have had early on, you're also delaying your impact further into the future. Most researchers on this topic agree that, all else being equal, it's better to put resources towards fixing the world's most pressing problems sooner rather than later. You might also give up on trying to have an impact in the meantime, and informal polls suggest the annual risk of this might be quite high. Finally, you gain *some* career capital from almost every reasonable option, so what matters is the *extra* career capital you get from the PhD compared to working as an AI safety engineer.

On the other hand, by making the right investments, it's possible to increase your impact a great deal, so the tradeoff can go either way. Overall, we're excited to see people take unusually good opportunities to gain career capital, especially those that open up specific paths that seem much higher impact.

Career capital is also more important earlier in your career on average, since you'll have more time to make use of your investments in future roles.

Further reading

- Our article on alternatives to consulting
- Our article on why it's important to have a financial safety net
- Read more of the arguments for and against delaying your impact in *Dynamic Public Good Provision under Time Preference Heterogeneity: Theory and Applications to Philanthropy* by Phil Trammel, or this shorter article by Toby Ord
- Our article on which jobs put you in the best long-term position
- Our article on all the evidence-based advice we found on how to be more successful in any job
- Holden Karnofsky on building aptitudes and kicking ass

Career exploration: when should you settle?

Click to read online

Suppose you've researched different career paths, and now need to make a choice:

1. 'Settle': commit to the path that seems best now.
2. Explore: try other paths with the hope of finding something even better.

What should you do?

Steve Jobs liked to say you should "never settle," but there's a real balance to be struck between exploring and committing.

Many hope to be able to find and commit to their career calling right away, but this is rarely possible because it's so hard to predict where you're going to succeed in the long term.

Rather, you should approach your career like a scientist doing experiments. This means you should be prepared to test out several paths, if possible.

While everyone would ideally do some career exploration, the interesting question is *how much* you should plan to explore, and how best to balance the costs of exploring with its upsides.

There's been plenty of research in decision science, computer science, and psychology that can help us answer this question. In this article, we combine these findings with what we've learned from advising people one-on-one, and summarise some of the bottom lines.

We'll argue that if you want a career that's not only satisfying but has a significant positive impact — our focus at 80,000 Hours — then the value of exploration is even higher.

Career exploration in a nutshell

- In addition to building career capital and having an impact, trying a job gives you 'information value' about which paths will be best for you in the long term.
- The paths with the highest information value are the ones that might be a lot better than your current best guess, but that you're very unsure about and can learn about relatively quickly.
- One career exploration strategy is to choose the path that would be highest-impact if it goes unusually well (in an 'upside scenario'), stick with it if it works out, and try something else if it doesn't. (Though make sure you're OK with this degree of risk and have a backup plan.)
- Another career exploration strategy is to plan to try out several paths, which could include 'wildcards' outside your experience, minimising the costs by carefully ordering your options. This strategy is most attractive right at the start of your career and for perhaps 2–8 years in total.
- In general, it's most important to explore early in your career, since you have more uncertainty at this stage and more time to benefit from what you learn, and the costs of trying different jobs and careers are lower.
- Be more open to quitting: people tend to stick in their current path too long even when it's not working out.

The information value of exploration

The typical 25- to 34-year-old changes jobs every three years, and changes are not uncommon later in life too.[45] Many successful people have tried several paths before finding one that worked out — Tony Blair even worked as a rock promoter before going into politics.

[45] "Median employee tenure was generally higher among older workers than younger ones. For example, the median tenure of workers ages 55 to 64 (10.4 years) was more than three times that of workers ages 25 to 34 years (3.0 years). A larger proportion of older workers than younger workers had 10 years or more of tenure. In January 2014, among workers ages 60 to 64, 58% were employed for at least 10 years with their current employer, compared with only 12% of those ages 30 to 34." Archived link, retrieved 24-Apr-2017.

Many people don't have the luxury of exploring different paths, but if you do, taking that time can easily be worth it. For instance, if exploring meant you could find a career path that's twice as good as your current best guess, it would theoretically be worth spending up to half of your career searching for that path.

We've argued that some career paths open to you will do 10 or 100 times as much to help the world as others, and so even if you spent 10 years searching for a path like this, it would be worth it — in your remaining 30 years of work, you'd achieve as much as would have taken 300 in your previous path. It's often possible to find paths that are more enjoyable or give you better skills too.

We can analyse the value of exploration by using a concept from decision science — 'information value.'

The information value of trying a job is what you learn about which longer-term paths are best. The value depends on the chance that you discover a path that's better than your current best guess, weighted by how much better that new path is (based on whatever you value).

When comparing jobs you might take during your job search, you care about both:

- The immediate impact of and satisfaction you gain from the job
- How well it sets you up to find better opportunities in the future

We can now see that the latter consists not only of the career capital you gain from the job (such as transferable skills, professional network, and reputation), but also the information value it gives you. We could say:

> *Contribution to long-term impact of job = immediate impact + career capital + information value*

The best job for you is the one that delivers the most value across all three factors, plus the personal satisfaction you gain from it.

A job could be a 'failure' in the sense that you didn't have any impact, or hated it and quit. But if you're able to eliminate a long-term path as a result of this experience, that can still be valuable — you've just saved yourself years of wasted effort.

To sum up: each job you take is also an experiment to help you learn about what's best for you in the long term.

Amanda switched from philosophy to AI policy at age 29. While doing that, she also wrote about how the 'information value' of testing different career options can be one of the most significant considerations when working out how to do good. Read more

Which jobs should you take to explore?

The paths that are best to try out (i.e. have the most information value) are those that:

1. Might be *a lot better* than your current best guess: if a path isn't much better than your current best guess, then even if your test goes well, you haven't gained much.
2. You're uncertain about: if you already know it's better, then you should just take it now.
3. Could let you reduce your uncertainty relatively easily: if trying the path wouldn't tell you much about your fit, then there's also not much to be gained from testing it — better just to take your best guess now.

Can you think of a career path that meets all three of these conditions? If so, there's a pretty good case for trying it out.

We often advise people who feel very uncertain about which paths are best for them, but are tempted to try out a corporate job, like accounting or consulting. While this *might* make sense because of the helpful career capital, it often doesn't have much information value — these people usually don't expect that the corporate path is their best long-term path, and they're normally pretty certain about that.

Tradeoffs of career exploration

The temptation to do the corporate job, however, does illustrate a real tradeoff: your best job for information value might not be the best for career capital, impact, or satisfaction.

Trying out a different career path not only takes years, but the costs can be even higher than they first seem.

Some competitive careers have an 'elevator' structure, where if you get off the standard track, it's hard to re-enter — and even if you *can* re-enter, it's hard to make it to the top. This can make it higher-risk to try out alternative paths early on, since by leaving you might miss your chance of outsized success.

Academia is the most common example of this among our readers. Going from a PhD to a postdoc, and then into a permanent academic position, is very competitive, and it's unlikely you'll succeed if you don't focus the vast majority of your effort on research. This makes it hard to re-enter academia if you try something else after your PhD.[46]

Exploring also *delays* your impact, which could be a significant cost if you think there are especially urgent opportunities to do good right now that won't be around in the future. For example, if you're trying to shape fast-moving technologies, that will be harder to influence as time goes on.

But we've also argued above that the potential gains to exploration are very large too, especially if you want to have an impact. Going from 'very good' to 'excellent' might only increase your own wellbeing a little bit, but it could mean having 10 times as much impact.

So, both the benefits and costs of exploration can be large. What strategy should you take to balance these?

Career exploration strategies

This question has been studied in computer science, as part of research into topics like the 'explore-exploit tradeoff' and 'optimal stopping' (of which the 'secretary problem' is one example). We discuss what we might learn from this research in-depth in our interview with Brian Christian, author of *Algorithms to Live By: The Computer Science of Human Decisions*.

The lesson we take from this research is that there are two exploration strategies that make sense practically. The first we call the upside option strategy, and is usually our first recommendation. The second involves trying out several paths early career, and then picking your best guess after that.

1. Pursue 'upside options': a rational reason to aim high

As we saw above, information value is higher the better an option might be, and the more uncertain you are about it. This favours taking *long shots* — paths that might be outstanding if they go well, but have a good chance of not working out.

[46] Though it's easier in some subjects like computer science or economics, and it's fine to take a break before your PhD.

Here's a concrete process for doing that:

1. Rank the longer-term paths open to you in terms of upside — how much impact you would have in them if you performed *unusually well* compared to your best guess? (To be more precise, you could imagine how each path would look in the top 10% of scenarios.) We call the career options that seem best in these 'upside scenarios' your 'upside paths.'
2. Try out one of your most promising upside paths.
3. If you find it's going well, then continue.
4. If you're not on track to hit the upside scenario, switch into the next-best upside path.

This strategy is similar to the 'upper confidence' algorithm we discussed with Brian.

It's attractive because there's an asymmetry: in the good case, you've found an amazing new career path, which you can continue with for many years; if it doesn't work out, you can switch to something else relatively quickly. The costs of spending a few months or even years trying out a path are often low relative to the huge benefits if it turns out well.

Taking this strategy is also attractive because if you're trying an elevator career, you can stick with it until you've decided you're not likely to have outsized success, which is often where most of the impact would come from. (There are also several other reasons we think it's good to aim high.)

However, there's an important caveat to keep in mind: the asymmetry argument only works if the downsides of exploration are capped:

- First, you have to make sure that you're in a sufficiently robust position — personally, professionally, and financially — that an experiment not working out won't do you considerable long-term harm. That means making sure you have a backup plan and remain able to switch into something else if it doesn't work out. We discuss how to do this in our career planning process.
- Second, if you're focused on impact, then in certain areas — 'fragile fields' — it's easy to make things much worse than you found them. If you're doing something high-stakes in one of these fields, it's important to modify your plan to reduce the chance of setting back the field before pursuing upsides. We cover this in our article on accidental harm.

Finally, note that what counts as an 'upside scenario' should depend on how much you want to prioritise exploration, which depends on how much you prioritise impact and how early you are in your career. For instance, if you're right at the start of your career, you might want to aim at what's best based on the top 5% of scenarios; if none work out after a while, then reduce that to the top 20% of scenarios — and eventually focus on the most likely scenarios.

2. Plan to try out several paths through careful ordering

The second career exploration strategy is to plan to try out several paths, and then decide later which is best. If you order your options carefully, it's often possible to try several paths with minimal costs, then decide later which is best.

This strategy is similar to solutions to the 'secretary problem' in computer science. The standard set up of the secretary problem seems to overstate how much you should explore, but spending 5–20% of the total length of your career with exploration as a top priority seems reasonable for people focused on impact, and that would be 2–8 years.[47]

One common pattern is to do a gap year before college, then try several internships while at college, then do something more unusual and risky for 1–2 years after graduating. If the unusual option goes well, you can continue, and otherwise you could try going to graduate school, which 'resets' you onto a more standard path. For instance, you could try an unusual startup or nonprofit project for a few years, and then do a PhD and continue with academia, or go to law school and continue into policy, and so on.

Later in your career, if you're genuinely unsure between two options, you might want to try the more 'reversible' one first. For instance, it's easier to move from business to nonprofits than vice versa.

While examples of people who specialised early, like Tiger Woods, are often salient, it doesn't seem *necessary* to specialise that early, and it's probably not even the norm. In the book *Range: Why Generalists Triumph in a Specialized World*, David Epstein argues that most people try several paths, and that athletes who try several sports before settling on one tend to be more successful — holding up Roger Federer as a foil to Tiger Woods.

Another recent paper in *Nature* found that 'hot streaks' among creatives and scientists tended to follow periods of exploring several areas. This is just a single paper, but at least suggests exploring several paths can be a good strategy in some cases.

Compared to the 'upside options' strategy, this strategy keeps you from committing to anything for years, so has significant costs. If you can identify an upside path now, trying it right away is usually better.

[47] Learn more in this analysis by Applied Divinity Studies.

Planning to try out several paths is most attractive when you're very uncertain, when you won't risk compromising your future in an 'elevator' career, and when you're early in your career (when the costs of exploration are lowest).

Jess – a case study in trying several paths

"80,000 Hours has nothing short of revolutionised the way I think about my career."

READ JESS'S STORY

When Jess graduated with a degree in maths and philosophy in 2012, she was interested in academia and leaned towards studying philosophy of mind, but was concerned that it would have little impact.

So the year after she graduated, she spent several months working in finance. She didn't think she'd enjoy it, and she turned out to be right, so she felt confident eliminating that option. She also spent several months trying different types of work — working in nonprofits and reading about different research areas.

Most importantly, she spoke to loads of people, especially in the areas of academia she was most interested in. This eventually led to her being offered to study a PhD in psychology, with a focus on how to improve decision-making by policymakers.

During her PhD, she did an internship at a leading evidence-based policy think tank, and started writing about psychology for an online newspaper. This meant that she was exploring the 'public intellectual' side of being an academic, and the option of going into policy.

At the end of her PhD, she can either continue in academia, or switch into policy or writing. She could also probably go back to finance or the nonprofit sector. Most importantly, she'll have a far better idea of which career options are best for her.

Consider including a wildcard

One drawback of both of the strategies above as written is that your best path might well be something you haven't even thought of yet.

This could suggest exploring and testing out options outside your normal experience, to give yourself the chance of uncovering something totally different. That might mean living in a very different culture, participating in different communities, or trying different sectors from the ones you already know (e.g. nonprofits, government, corporate).

For instance, I (Benjamin) went to learn Chinese in China before I went to university. I didn't have any specific ideas about how it would be useful, but I felt I learned a lot from the experience, and it turned out to be useful when I later worked with Brian Tse to create our resources for people working on Sino-Western coordination around emerging technologies.

This is similar to how in computer science, many exploration algorithms have a random element. Making a random move can help avoid settling into a 'local optimum.' While we wouldn't recommend literally picking randomly, the fact that even computer algorithms find randomness helpful illustrates the value of considering paths well off your standard track.

3. Explore within your job

An alternative approach is to take a job that lets you try out several areas by:

- Enabling you to practise many different skills. Jobs in small companies are often especially good on this front.
- Enabling you to work in a variety of industries, such as certain types of freelancing or consulting positions.
- Giving you the free time and energy to explore things outside of work.

Be more willing to quit

Another rule of thumb to consider: if you're on a path that's only going so-so, you should probably try something else. (If you can afford to.)

If the path is only going so-so, it's probably not an upside option, and so doesn't have much information value. It's probably also not a great fit.

Due to the sunk cost bias, we'd expect people to continue with their current path for too long. We'd also expect people to want to avoid the short-term costs of switching, and to be averse to leaping into an unknown new option.

This all suggests: if you're on the fence about quitting your job, you should quit.

This is exactly what an influential randomised study found. Steven Levitt recruited tens of thousands of participants who were deeply unsure about whether to make a big change in their life. After offering some advice on how to make hard choices, those who remained truly undecided were given the chance to flip a coin to settle the issue — 22,500 did so.

Levitt followed up with these participants two and six months later to ask whether they had actually made the change, and how happy they were one a scale of 1 to 10. It turned out that people who made a change on an important question gained 2.2 points of happiness out of 10. Of course, this is just one study, and we wouldn't be surprised if the effect were smaller on replication. However, it lines up with what we'd expect.

How much should you explore at different stages in your career?

It varies by person. If you're following the upside option strategy, and you luck into something great early on, you might never need to try anything else. Or, you may need to try over and over again to get a hit.

However, there are clear arguments that people should generally explore more earlier in their career. This is because when you're early on:

1. You have the most time remaining to take advantage of new options you discover, whereas if you find a new career path when you're 60, you can only benefit for a few years.
2. You have the least information about what's best, since you haven't tried many jobs.
3. The costs of exploring are lower: young people are expected (often *encouraged*) to change jobs more often, and there are low-cost opportunities to try things out (like internships), which aren't always available to older people.

This is the justification behind the common advice to be more ambitious and broaden your horizons while young. The younger you are, the more you can focus on information value — and the more you're focused on information value, the more ambitious the upside paths to aim for. As you get older, you'll focus more on 'exploiting' your best-guess choice rather than exploring to find something even better.

However, you'll keep learning about your fit and which options are best throughout your career, so information value remains important for everyone. A middle-aged person who has just quit a long corporate career would suddenly face more uncertainty and lower costs to experimentation than most people their age, and so might do well to go back into exploration mode for a period.

Conclusion: exploring step-by-step towards your best career

It can be intimidating to try something new, but it's often well worth your while. You don't need to figure out your best career right away, and it's not the norm. Most careers proceed via a series of steps, each of which is an experiment. The best you can do is have a good exploration strategy — one that minimises the costs of exploring while giving you the best possible chance of finding an even better path.

Further reading

- Our podcast with Brian Christian on what we can learn from computer science about how much to explore in your career.
- How long should you take deciding your career? on Applied Divinity Studies further investigates how to apply the secretary problem to career choice.
- If you're unsure whether to quit, probably quit.
- David Epstein on Mastery, Specialization, and Range, on EconTalk (the book itself doesn't add much to the arguments, though has more examples).
- How to Measure Anything by Douglas Hubbard has one of the best explanations of 'information value' that we've seen, and is a great book in general. Here is a summary.
- Four examples of value of information.
- Biases in career decisions and what to do about them.

How to balance impact and doing what you love

Click to read online

We think there's less tension between the two than is often supposed. Finding work you excel at and that helps others is fulfilling, and many of our readers say they've become happier in the process. Moreover, you'll have a greater impact if you find work you enjoy and that fits with your personal life, because you'll have a greater chance of excelling in the long term. So enjoying your work and having an impact are often mutually supportive goals.

This said, sometimes conflicts do arise. For instance, the higher-impact path may involve working harder than would be ideal for your happiness, or it can involve taking the risk of trying out several paths that don't go anywhere. How to handle these conflicts is a difficult issue.

We may live in a uniquely important time in history, with the opportunity to influence the development of new technologies that could impact the long-term future and reduce existential risks. We also have many other opportunities to help others a great deal with comparatively little cost to ourselves. This motivates some of our readers to make impartially doing good the main focus of their careers. Some philosophers, such as Peter Singer, have argued that we have a moral obligation to do so.

However, most of our readers see 'making a difference' in the way we've outlined as one among several important career goals, which may include other moral aims, supporting a family, or furthering other personal projects.

Whatever your views on this topic, we think it's important to take seriously the risk of burning out if one engages in too much self-sacrifice. Even if your only career goal were to make a difference, you should probably be aiming to contribute sustainably over your entire 40-year career. This means it's important to cultivate self-compassion and take a path in which you'll be motivated for the long term.

What's more, one of the biggest ways to have more impact is to inspire others to contribute, and this is much easier when you're enjoying your life and career.

One technique that can be helpful is setting a target for how much energy you want to invest in personal vs. altruistic goals. For instance, our co-founder Ben sees making a difference as the top goal for his career and forgoes 10% of his income. However, with the remaining 90% of his income, and most of his remaining non-work time, he does whatever makes him most personally happy. It's not obvious this is the best tradeoff, but having an explicit decision means he doesn't have to waste attention and emotional energy reassessing this choice every day, and can focus on the big picture.

Further reading

- The ingredients of a satisfying career and our research behind this.
- Famine, Affluence and Morality
- Here are some philosophical arguments against the view that we have a moral obligation to do good even when doing so involves substantial sacrifice: Satisficing Consequentialism and Morality and Reasonable Partiality. On the other side: Does Consequentialism Demand Too Much?
- Giving Gladly

Be more ambitious: a rational case for dreaming big (if you want to do good)

Click to read online

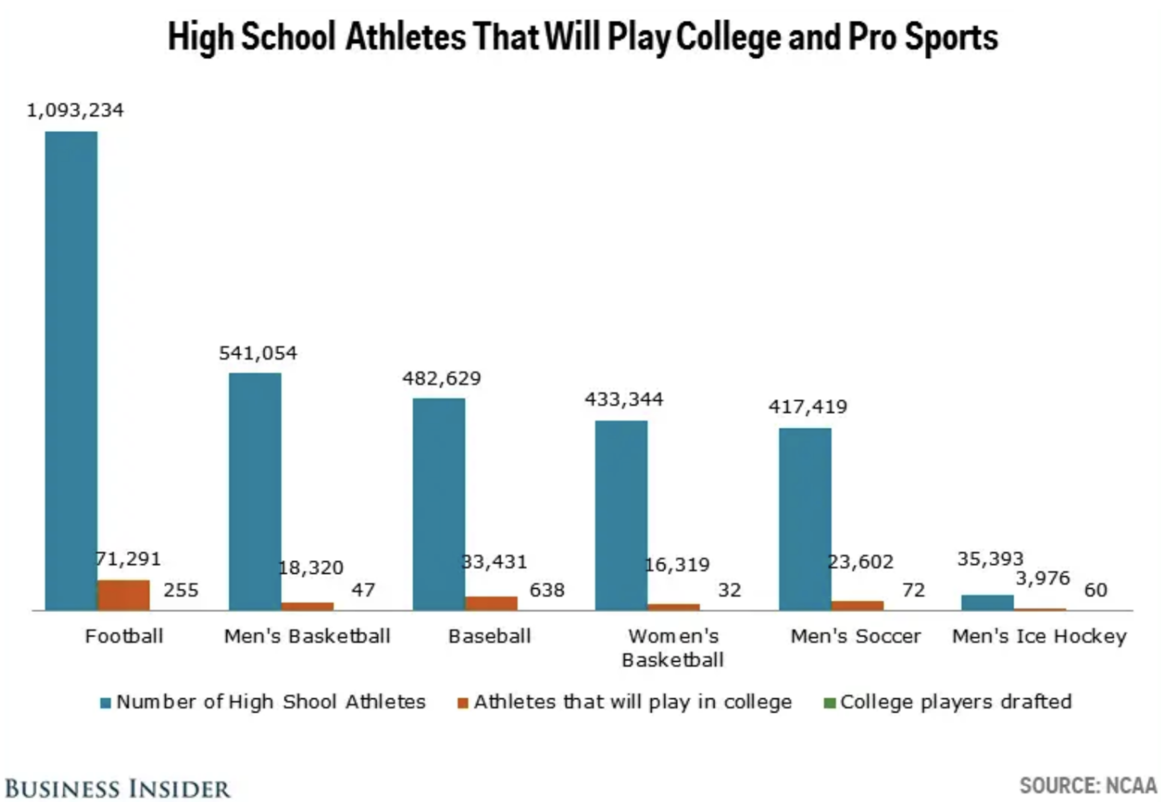
Self-help advice often encourages people to "dream big," "be more ambitious," or "shoot for the moon" — is that good advice?

Not always. When asked, more than 75% of Division I basketball players thought they would play professionally, but only 2% actually made it.[1] Whether or not the players in the survey were making a good bet, they overestimated their chances of success… by over 37 times.

This level of overconfidence is common, and means that "be more ambitious" may not always be the right advice. Some people even enjoy taking risks, which explains why they buy lottery tickets even though they lose money on average. Whether to be more ambitious depends on the domain and the person in question.

However, if your aim is to have a positive impact on the world, we think we can make a rational case for setting ambitious goals — based on the concepts in our key ideas series.

In short, our advice is to do as much as you can to set up your life so that you can afford to fail, and then aim as high as you can. As a slogan: limit downsides, target upsides.



The fraction of high school athletes who will go pro is tiny. Even among Division 1 college athletes, 44–76% believe they will go pro (depending on the sport), but typically under 2% actually make it — the odds are best in baseball.

In a nutshell: why to be more ambitious

If you want to do good, here are four reasons to be more ambitious:

1. If you care about the number of people you help, it can be worth betting on a small probability of helping a huge number (unlike with personal goals, where most things have diminishing value).
2. The wide variation in how much good different career paths can do means that low-probability, high-upside scenarios can be the biggest driver of your impact.
3. Aiming high has more information value, since you give yourself the chance of being positively surprised.
4. Other actors are risk-averse, so you'll face less competition.

In order to free yourself up to be ambitious, first limit downsides:

- Modify or eliminate options that might have a serious negative impact on you or the world, or prevent you from trying again.
- Make sure you have a backup plan.
- Put yourself in a better position to take risks over time by investing in your financial security, skills, and mental & physical health.

What do we mean by being more ambitious?

We mean you should aim high. More concretely we mean you should:

1. Make a list of longer-term career paths you could aim towards.
2. Think about how much positive impact you'd have if each path goes really well (what we call an 'upside scenario').
3. Think about what will happen if the path goes badly. Modify or eliminate any options that might have a big negative impact — either on your life or on the world.
4. Then, to choose between the remaining paths, seriously consider pursuing the one with the best outcome in the upside scenario.

To be more precise, we can define the upside scenario as the top 5% or top 10% of possible outcomes. But even when you have no idea what the probabilities are, you can still broadly target paths with high upsides. (You can also apply this rule of thumb at the level of setting goals for specific projects, though our focus here is on longer-term career paths.)

Aiming high means looking for paths that might turn out really well, even if they seem wacky, or there's a good chance they won't work out.

Four reasons to be more ambitious

When we advise people on their impact, they often want to feel confident they've done 'some' good with their career, rather than bet on a small chance of achieving much more. This is understandable, because it's satisfying to know that you've achieved something rather than nothing.

But if your aim is to help people, rather than just feel satisfied, it could be a big mistake. By being open to long shots, ambitious people can greatly increase their expected impact — even accounting for the chance their projects fail.

Here are four reasons why people who want to do good should be *more* open to taking risks than the norm.

1. A small chance of achieving a lot can still be worth a great deal (& much more than in your personal life)

There are realistic ways you can make a contribution to some of the most important and neglected issues of our time.

More than that, there's a chance you could achieve something amazing. Maybe you can start an important new charity, or run a media campaign that shifts people's views of a crucial issue. Maybe you could even win a Nobel Prize, be elected to Congress, become prime minister, found a 'megaproject' nonprofit, or become a billionaire. If you go in with a firm commitment to doing good, any of these wild achievements could allow you to have an outsized impact.

These kinds of outcomes are all unlikely of course. But they might be more likely than you think.

Several thousand people have changed careers based on our advice, and one of them has already become a billionaire who has pledged his wealth to helping solve global problems, and is also probably the world's richest person under 30.

Sam Bankman-Fried is extremely talented and comes from a privileged background, but one of his most notable traits (and also critical to his success) is that he aims high. Sam probably took the idea of becoming a billionaire more seriously than 99% of our readers. What would happen if you took the possibility of wild success seriously?

"80,000 Hours helped me think more critically about my career choice, which has had a significant impact on where I am today."

READ SAM'S STORY

> *One of the general themes that I've become more and more convinced by over time is that all the expected value is in the upside tails, and not in the median outcomes usually. You should take that seriously, and it implies weird things. Like that often the right path is the one that very well might fail. You should be asking yourself the whole time: what is upside? What might it look like? Where is it?*
>
> *And I think that as the world speeds up and gets wackier, that becomes more and more important. Sometimes things that sound crazy and unlikely, might be unlikely, but maybe not so unlikely, that they're not super valuable in expectation.* – Invest Like the Best Podcast

Even if the chance of an amazing outcome is very low, it could still be worth betting on. Why?

From the perspective of making the world better, helping two people is twice as good as helping one.

Taking this idea further, a 10% chance of saving 200 lives is better than a 90% chance of saving 10. This follows from the idea of expected value.[48]

If thousands of our readers all pursued projects with a 10% chance of success, there would be hundreds of successes, even if most don't work out.

These are toy examples, and in reality we'll never know the probabilities or outcomes with much confidence. But it illustrates that it can be worth setting ambitious goals and betting on comparatively unlikely outcomes — if the odds and upsides are high enough.

This is much less true in your personal life. Gaining 10 new friends isn't twice as good as gaining five, since what we most care about is feeling like we have a community of some kind. Most people would prefer $10 million with certainty compared to a 20% chance of $100 million. This is because money (and most other things) have diminishing value for individuals. So in your personal life, long-shot gambles are much less attractive — aim for 'good enough' rather than maximising the potential upside.

---

[48] Everything we do has uncertain outcomes, but if we want to make a bigger difference we should do what has the best expected outcome. This means multiplying the value of the outcome by the probability of it occurring. 200 lives saved x 10% likelihood = 20 lives saved in expectation. 10 lives saved x 90% probability = 9 lives saved in expectation. Using math in doing good might seem weird, but when we're dealing with uncertain outcomes (which we always are), we need it to choose between actions.

Psychologically, most people are also loss-averse — losing a friend is a lot more painful than gaining one is joyful — which makes it even more important to avoid big risks.

None of this applies to making the world a better place, where helping twice the number of others, or helping them twice as much, really *is* twice as good. If you've saved one life already, that doesn't make it any less valuable to save another. So do-gooders should be more open to taking risks.

This effect becomes more extreme the higher you aim. From a personal point of view, gaining $10 million with certainty is a lot more attractive than a 2% chance of gaining $10 billion. But if you're going to donate the money, the second approach is about 20 times higher impact. Likewise, most people would prefer to have a job that's 'pretty satisfying' with near certainty, rather than a 2% chance of making a scientific breakthrough.

(That's not to say you should be risk-neutral. You shouldn't be completely risk neutral because there are still diminishing returns to altruistic resources at large enough scales. This is especially true if you're investing, since your investments will probably be correlated with other donors. But it is justified to take a lot more risk than normal.)

In short, 'satisfice' your personal goals, but maximise your impact.

2. Upside scenarios can be where most of the expected impact is from

In the world of doing good, there's a large skew in outcomes. Most importantly, we've seen that in many fields, especially those like research or entrepreneurship, the most successful people are often responsible for a significant fraction of the total impact. (Just as Open Philanthropy and Sam are likely to be responsible for a significant fraction of donations in effective altruism.)

This means that the expected value of entering the field is significantly driven by the value of the upside scenario (adjusted for how likely it is). So, 'having the most impact' boils down to 'focus on the (non-crazy) scenario with the most upside.'

This won't always be true, but it's at least worth seriously thinking about what great outcomes might be possible, on the chance it is possible for you.

In our article on effective solutions, we tell the story of how Sophie Rose tried to make a big contribution to ending COVID-19 early by advocating for human challenge trials. Although she wasn't able to succeed quite in time, her work will better prepare us for the next pandemic.

More speculatively, focusing on upsides might be becoming more important as technological advancement means that extreme changes are happening more and more often.

One dizzying implication of this way of thinking is that from the perspective of your impact, if one scenario is far higher impact than the others, it can sometimes be ideal to act *as if* it's going to happen, and basically ignore the other scenarios.

For instance, if the 21st century is the most important century in history, then our actions could have truly massive significance. If that's not the case, the 21st century won't matter nearly as much from a historical perspective (in which case it doesn't matter as much what we do). Because our actions matter far more in the first scenario than the second, we should act as if we know the first scenario is true.

3. You might surprise yourself

If you aim high, you might be positively surprised: maybe you can actually achieve the upside scenario, and you've discovered an amazing path.

If, however, you discover the upside scenario is not going to happen, you can probably switch to something else without great costs.

In other words, there's an asymmetry: by aiming high, you might find an amazing career path, while if it doesn't work out, you're probably in a similar situation to before. A lack of ambition would cut you off from this possibility.

You can read more about this in our article on exploration, where we show that you stand to learn the most from pursuing paths that *might* be amazing, but where you're also really uncertain how they'll turn out. The greater the uncertainty, the more you'll learn by trying it.

We've worked with lots of people who applied to a new job way out of their comfort zone, thinking it was a long shot, and then not only went on to land the position, but also excel in it.

4. Low-probability bets are neglected

Is the world of doing good more dominated by people who are overconfident (like would-be professional basketball players) or people who are risk-averse? Our impression is that it's the latter. If so, this means high-upside paths will be relatively neglected — perhaps especially at the very highest end.

Why do we have this impression? One reason is that many efforts to do good are done by governments and nonprofits, and their incentives make it hard for them to take low-probability, high-upside bets.

Suppose a government bureaucrat can fund a programme that has a 10% chance of an amazing outcome, like speeding up vaccines for COVID-19 in 2020. After funding five projects that don't work, this bureaucrat will probably lose their job, even if the expected value of each project was very high. On the other hand, if the bet pays off, they'll get few rewards — maybe a bit of praise from their colleagues, or a modest raise, if they're lucky. These incentives make people risk-averse.

We've seen some empirical evidence of this. In one study, the Howard Hughes Medical Institute took a more risk-tolerant approach to funding medical research and seemed to get better results than the US National Institutes of Health — the more conservative government agency.

Carl Shulman discusses these dynamics on our podcast. Open Philanthropy also argues that in the world of philanthropy, higher-risk bets are often better.

In other words, since most people don't aim high, as you aim higher, you face less and less competition. This means the odds of success decline more slowly than you might expect. For example, while it might be harder to found a nonprofit with a budget of $100 million than $10 million, it's not obvious it's 10 times harder.

Limiting downside risks

We've given the arguments for aiming high, but you can only set truly ambitious goals after you've limited the downsides, so we'll also talk briefly about why and how to do that — setting yourself up to be as ambitious as possible.

Why limit downsides first?

If you aim high, there's a good chance you won't succeed. You'll be better able to maximise your long-term impact if you're able to "stay in the game" and keep trying until you succeed. So, it's important to minimise any risks of outcomes that might prevent you from trying again, like damaging your mental or physical health, or ruining your reputation.

We've also seen that it's perfectly reasonable to be risk- and loss-averse about your personal life. This means it's important to avoid risks that might make you a lot less happy, or embark on paths with unacceptable downsides, like not being able to support someone who is financially dependent on you.

Finally, in the world of doing good, it's sometimes possible to make things a lot worse than they were before. This is different from, say, the world of startups or investing, when normally the worst-case scenario is that entrepreneurs lose your original investment, and the risk is naturally capped.

The argument above to 'focus on upsides' only works when the potential for making things better is a lot higher than making them worse. If you're considering a path or project that might have a big negative impact (rather than merely 'fail' and not achieve much), then you need to carefully weigh the upsides and downsides. Most of the time, we'd recommend simply avoiding projects like this, and then focusing on the biggest upsides among your remaining options.

How to limit downsides

Even if you can't easily estimate how likely risks are to materialise, you can often do a lot to limit them, freeing yourself up to focus on upsides.

Over time, you can aim to set up your life to make yourself more able to take risks. Some of the most important steps you can take include:

- Building up your financial security. If you're at constant risk of failing to make your rent, that's a serious downside you can't discount.
- Looking after your physical and mental health and important relationships, so that your lifestyle is sustainable.
- Building valuable career capital that gives you backup options, e.g. through building skills or finding good mentors.

When comparing different career paths, here are some tips:

1. Consider 'downside scenarios' for each of the paths you're considering. What might happen in the worst 10% of scenarios?
2. Look for risks that are really serious. It's easy to have a vague sense that you might 'fail' by embarking on an ambitious path, but what would failure actually be like? The risks to be most concerned about are those that could prevent you from trying again, or that could make your life a lot worse. You might find that when you think about what would actually happen if you failed, your life would still be fine. For example, if you apply for a grant for an ambitious project and don't get it, you will have just lost a bit of time.
3. If you identify a serious risk of pursuing some option, see if you can modify the option to reduce that risk. Many entrepreneurs like Bill Gates are famous for dropping out of college, which makes them look like risk-takers. But besides the security provided by his upper middle-class background, Gates also made sure he had the option to return to Harvard if his startup failed. By modifying the option, starting Microsoft didn't involve much risk at all. Often the most useful step you can do here is to have a good backup plan, and this is part of our planning process.
4. If you can't modify the path to reduce the risk to an acceptable level, eliminate that option and try something else.
5. Check with your gut. If you feel uneasy about embarking on a path even after taking the steps above, there may be a risk you haven't realised yet. Negative emotions can be a sign to keep investigating to figure out what's behind them.

Conclusion: it's better to be too ambitious than not ambitious enough

We advise people who are overconfident, as well as people who are underconfident. But if your aim is to have an impact, underconfidence seems like the bigger danger. It's better to aim a little too high than too low.

But ambitious people do not need to be irrational. You don't need to convince yourself that success is guaranteed. To be worth betting on, you just need to believe that:

- Success is possible
- Your downsides are limited
- The expected value of pursuing the path is high

If you've found a path that might be amazing, make a backup plan and give it a go. It may not work out, but it might be the best thing you ever decide to do.

If you're inspired to start setting goals and update your career plan right now, take a look at our planning process.

Further reading

- How much risk to take?
- Expected value: how can we know what makes a difference when uncertain?
- Black swan farming by Paul Graham
- Optimism for realists and the optimism heuristic by Danny Hernandez
- Julia Galef argues you don't need to be irrationally overconfident to succeed in *The Scout Mindset*. See this interview with her by Noah Smith.
- When should altruists be financially risk averse? & Should altruists use leverage? by Brian Tomasik
- Risk aversion and investment for altruists by Paul Christiano

3 key career stages

Click to read online

If you want to have an impact, the aim is to find a job that has the potential to make a big contribution to a pressing problem, and that's a good fit for you. But how can you find a job like that?

In the strategy section of our key ideas series, we discuss the value of exploration and career capital, as well as many other ideas, like why to be more ambitious. Here we sum them up into a simple career strategy.

Three career stages

How to find a great job in a sentence: get good at something that lets you effectively contribute to pressing global problems.

To expand:

First, make some best guesses about which longer-term roles seem best to aim towards, both in terms of impact and career capital. We've compiled a list of high-impact options to help you get started, as well as advice on comparing them. Your answer will only be a guess because it's hard to predict where you'll succeed in the long term. We advocate taking an iterative approach, updating your best guess every 1–3 years.

Once you have some best-guess longer-term roles, you can roughly follow these three stages:

1. Explore: take low-cost ways to learn about and test out promising longer-term roles, until you feel ready to bet on one for a few years. Exploration is most likely to be the top priority ages 18–24.
2. Invest: take a bet on a longer-term path that could go really well (i.e. seek upsides), by building the career capital that will most accelerate you in your chosen path. In case it doesn't work out, have a backup plan (i.e. limit downsides). (Age 25–35.)
3. Deploy: use the strengths and career capital you've built to help implement the most effective solutions to the most pressing problems at the time. (Age 36 and up.)

These stages are just about what to emphasise at each time. In reality, it's always valuable to gain information about which paths are best, invest in your career capital, and try to have an impact right away — and you might try to serve any of these priorities at any stage in your career.

The stages last different amounts of time for different people. You might find a path worth betting on straight after university, and if it goes well, never look back. For example, Kuhan realised that his position as a student at Stanford gave him a great platform to spread important ideas. He helped to start the Stanford Existential Risk Initiative, which has helped hundreds of people learn about existential risks. Kuhan skipped straight to deployment — which worked out well for him — but most people probably need to try many different things before they find a path that works.

Your emphasis might also move back and forth over time. For instance, a 40-year-old who decides to make a dramatic career change might go back into exploration mode for several years. It all depends on your specific situation.

What stage are you in?

Besides your age, here are some questions to help you decide where to focus:

- How uncertain are you about which roles are the best fit? The more uncertain you are, the more you should focus on exploration. But don't wait too long before taking a bet (a common mistake). You just need to find a role that's worth betting on, rather than one you're confident in. If you have a good backup plan, the downsides of taking a gamble are low, and just trying a path is often the best way to learn about your fit with it.
- Are you learning? It's often worth staying in investment mode as long as you're learning and becoming significantly more productive.
- How urgent are opportunities to contribute? Taking time to invest needs to be weighted against the urgency of contributing to your chosen problems. If there are opportunities to have a big impact today that are much better than those you expect to find in the future, that can push you into deployment mode right away. This could be due to the time being unusually pivotal (e.g. a pandemic is starting) or because you've found an unusually good opportunity you might not have in the future (e.g. Kuhan from above).
- How uncertain are you about which global problems will be the most pressing in the future? The more uncertain you are about which global problems will be most pressing in the future, the more you should focus on finding roles that can be applied to many problems (e.g. manager, journalist, civil servant), rather than narrow or specialist roles (e.g. expert in market stabilization for developing economies). Similarly, the more uncertain you are, the more you should focus on building *transferable* rather than specialist career capital when investing. The same arguments apply if you're uncertain about what your *personal* preferences will be in the future.

Why not to "keep your options open"

People who feel uncertain about their career often try to "keep their options open." There's truth in this advice (as we'll cover), but when people try to "keep their options open" in practice, it often ends up being counterproductive. In particular, it often leads people to:

- Defer thinking about which paths are actually best, rather than doing more research to reduce their uncertainty.
- Stay in paths they know they don't want to do long-term "to build career capital" — even when those paths aren't likely the best route towards their top longer-term paths. (We often see people doing this with roles in professional services and consulting).

- Pursue a middle-of-the-road path, rather than commit to something that might be great in terms of career capital or impact (while bearing in mind they could change paths if it doesn't work).

Of course, all else equal, it's better to have more options rather than less. For example, you may be genuinely uncertain between two paths. If path A would let you switch into path B, but it would be hard to move from path B to path A, that's a good reason to start with path A (and then switch into B if it doesn't work out).

But for most purposes, we think other rules of thumb are more important to follow than simply "keeping options open."

The more fundamental rule of thumb is to "seek upsides and limit downsides." Having a great option potentially increases your upsides (if you take the option and it's indeed great) while having limited downsides (you don't *need* to take the option if you don't want to). But there are many other ways to seek upsides and limit downsides.

In particular, we usually find it's more productive to do something more unusual and ambitious (aiming high), but with a backup plan (to limit your downsides).

Another more fundamental rule of thumb is to try to gain the best career capital you can — that usually does more to increase your options in the future than trying to "not close any doors."

In sum, we recommend the following three steps as a more productive way to approach career strategy than "keeping your options open":

1. First, if you're uncertain about what's best longer term, try to research your options to reduce that uncertainty and make a best guess.
2. Then try out high-upside paths — whether those updates are in terms of impact or career capital — (but with a backup plan to limit downsides).
3. Finally, you don't need to have it all figured out right away: update your plan every few years with what you've learned.

It can be unpleasant to face this uncertainty, but these steps will give you the best chance of managing it, and of finding the best possible path for you.

What's next

Now read the final article in the key ideas series, which sums up the most important ideas we've covered, and why you might be able to find a career with perhaps 100 times as much impact.

Then, you can explore all our best resources on career planning, and start generating (and narrowing down) the concrete options that might fit you best.

Other key articles on career strategy

- Doing good together: coordinating as a community
- How to be more successful: research-backed ways to invest in your personal development within your current job
- How can you make the most of your time at college?
- How to avoid accidentally doing harm
- How much risk should you take?
- Alternatives to prestigious corporate jobs for career capital
- What makes for a satisfying job?

8. Conclusion: how much do careers differ in impact?

Why some of your career options probably have 100x the impact of others

Click to read online

We believe that some of the career paths open to you likely have over 100 times more positive impact than other paths you might take.

Why? In our key ideas series, we've shown that you can have more impact by:

1. Finding a **bigger and/or more neglected problem**
2. Finding a more **effective solution**
3. Finding a **path with more leverage**
4. Finding work that **fits you better**

We've also shown that there are big differences for each factor:

- Some problems seem hundreds of times more neglected relative to their scale than others.
- Some solutions bring about 10-100 times more progress on the problem per year of work.
- Some paths let you have many times more leverage on those solutions.
- You can have many times more impact in a path that's a good fit.

On top of that, you can further increase your impact by having a good career strategy, such as by striking the right balance between investing in yourself and having an impact right away.

But also note that the differences *multiply together*, rather than merely add up.

For instance, if you can find a problem where additional resources are twice as effective, and can direct twice as many resources to that problem through greater leverage, then you'll have four times as much impact.

If you can find a path that's twice as good on each dimension, it would be 16 times higher impact in total.

And we've shown that despite the huge uncertainties involved, it often seems possible to find an option that's 10 times better within each factor. If you could find a path that's

10 times better, then when multiplied together, the differences across all factors could be 10,000 times.

In practice, it's normally not possible to find an option that's better on every dimension. For instance, by changing the problem area you're working on, you might have less confidence in your fit.

There are also some reasons to be sceptical of claims of outsized differences that we've covered elsewhere in the series, such as regression to the mean and epistemic humility.

All considered, however, we think it's often achievable to find a path that is 10 times more impactful than what you're currently focused on, and sometimes over 100 times more impactful.

It's easy to gloss over the significance of a 100 times difference, but let's appreciate for a moment just how much it matters. It could mean saving 100 times more lives, reducing carbon emissions 100 times as much, or making 100 times more progress reducing the biggest risks facing humanity.

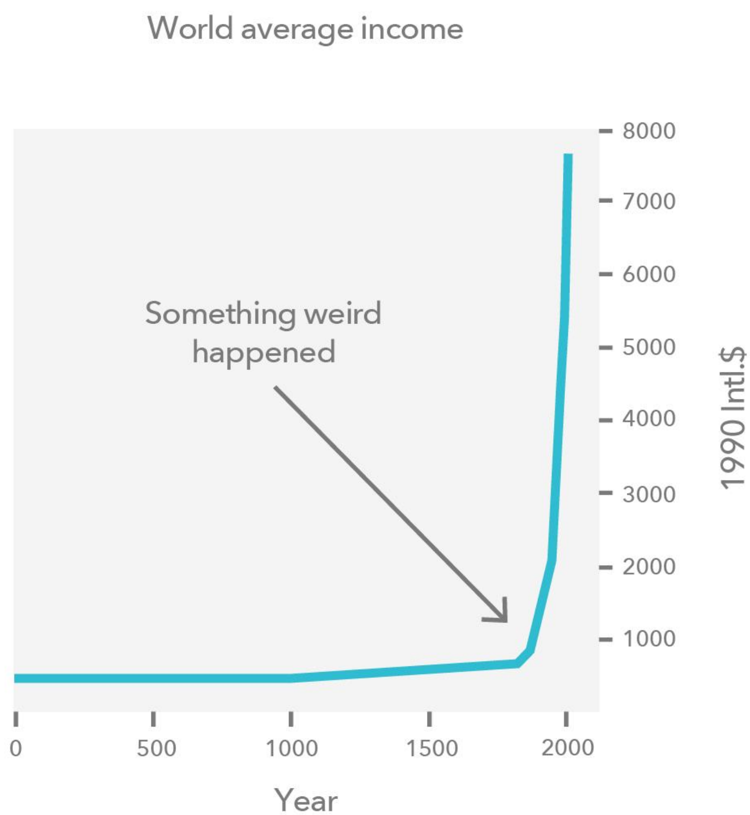
These differences are not the only ethically relevant factors, and everyone has priorities in life besides moral ones, but they do really matter.

Another implication is that if it's possible to find an option that's 100 times higher impact than your current best guess, then 10 years in that path would achieve what could have otherwise taken people like you *1,000 years*. You could then spend the next 30 years on a beach doing whatever you like, and still have done far more to help others.

This shows that it could easily be possible to find a path that's both higher-impact and more personally satisfying than your current trajectory.

How is it possible that such big differences in impact exist?

One reason is the massive economic and technological bounty of the Industrial Revolution, which means that today, many ordinary citizens of rich countries have what would have been kinglike wealth and power in previous centuries.



And advancing technology may make the next century one of the most important in history.

Our generation can wreck the climate for thousands of years, or we can build a sustainable economy. We can continue to expand factory farming, or we can eradicate it. We can allow new technologies like nuclear weapons to end civilisation, or we can usher in a future better than we can imagine — and be good stewards for all future generations.

Our aim is to help people like you understand this new power. If you have the good fortune to have options about how you spend your career, you can help change the course of history on these vital issues.

This is not an easy path, but it is a worthwhile one.

Finding the best path for you requires exploring, investing in yourself, being attentive to risk, and making hard tradeoffs.

We've often felt racked with uncertainty about what to do, and overwhelmed at the scale of the issues. But we've also found a meaning and satisfaction in our efforts, especially as more and more people have united around them.

We still have a lot to learn, but we hope that by sharing what we've learned so far we can help you avoid the mistakes we've made, and speed you along your path to an impactful career.

You have 80,000 hours in your career. Make them count.

What's next: make your new career plan

Now that you've read the series, here are two ways we can help you update your career plan, and put these ideas into action:

1. Speak to our team one-on-one

If you're interested in working on one of our top problem areas, our advising team might be able to speak with you one-on-one. They can help you consider your options, connect with others working on these issues, and possibly even help you find jobs or funding opportunities.

2. Use our planning course to write your plan

If you want to think more about your plan first, sign up for our eight-week career planning course. It takes everything we've learned about career planning and turns it into a series of tips, prompts, and resources to help you clarify your longer-term goals and turn them into actionable next steps. It's designed to be useful no matter which problems and career paths you want to focus on.

You can see everything online, or sign up to receive the course through a weekly email.