CHAPTER 10

# What to Do

## Backs to the Future

In the English language, the future is ahead of us and the past is behind. We might say that we must prepare for what lies before us and that we should not worry about what is behind us, or that we are facing a precarious future, or that Mary Wollstonecraft was a thinker ahead of her time. It turns out that this metaphorical mapping is near universal across cultures: as far as we know, every language in the world represents the future as being in front of us and the past as being behind, with just a handful of exceptions.[1]

The best-studied exception is the Aymara language. The Aymara are an Indigenous nation, comprising nearly two million people, who live in Bolivia, northern Chile, Argentina, and Peru.[2] Their traditional dress is brightly coloured, and their flag resembles technicolour glitch art. In the Aymara language, the future is behind us and the past is in front of us. So, for example, the phrase *nayra mara* is composed of the word for "front" (which also can refer to "eye" or "sight") and the word for "year," which means "last year." *Nayra pacha* literally means "front time" but refers to a "past time." To say "from now on," one says *akata qhiparu*, literally, "this from behind towards," and to refer to a "future day" one says *qhipüru*, literally, "behind day."

This conceptual metaphor is not restricted to Aymara speakers' choice of words. When referring to an event in the future, an Aymara speaker might point their thumb over their shoulder. This effect even persists when native Aymara speakers talk in a second language like Andean Spanish.

Almost all languages represent the future as ahead of us because when we walk or run, we both travel through time and travel forward through

space. In the Aymara language, the more important feature of time is what we know and what we don't. We can *see* the present and the past; they are laid out before us. We can therefore have direct knowledge of them in a way we can't know the future—anything we know or believe about the future is based on inference from what we have experienced in the present or the past.[3] The implicit philosophy is that, when making plans for the future, we should take much the same attitude as if we were walking backwards into unknown terrain.

This metaphor is an appropriate way to think about our journey into the future. Over the last nine chapters, I hope I've shown that it's possible both to think clearly about the future and to help steer it in a better direction. But I'm not claiming it's easy. At best, I've given a quick over-the-shoulder glance at the future that lies behind us. There is still so much we don't know.

Even over the course of writing this book, I've changed my mind on a number of crucial issues. I take historical contingency, and especially the contingency of values, much more seriously than I did a few years ago. I'm far more worried about the longterm impacts of technological stagnation than I was even last year. Over time, I became reassured about civilisation's resilience in the face of major catastrophes and then disheartened by the possibility that we might deplete easily accessible fossil fuels in the future, which could make civilisational recovery more difficult.

We are often in a position of deep uncertainty with respect to the future for several reasons. First, for some issues, there are strong considerations on both sides, and I just don't know how they should be weighed against each other. This is true for many strategic issues around artificial intelligence. For example: Is it good or bad to accelerate AI development? On the one hand, slowing down AI development would give us more time to prepare for the development of artificial general intelligence. On the other hand, speeding it up could help reduce the risk of technological stagnation. On this issue, it's not merely that taking the wrong action could make your efforts futile. The wrong action could be disastrous.

The thorniness of these issues isn't helped by the considerable disagreement among experts. Recently, seventy-five researchers at leading organisations in AI safety and governance were asked, "Assuming that there will be an existential catastrophe as a result of AI, what do you think will be the

cause?"[4] The respondents could give one of six answers: the first was a scenario in which a single AI system quickly takes over the world, as described in Nick Bostrom's *Superintelligence*; second and third were AI-takeover scenarios involving many AI systems that improve more gradually; the fourth was that AI would exacerbate risk from war; the fifth was that AI would be misused by people (as I described at length in Chapter 4); and the sixth was "other."

The typical respondent put a similar probability across the first five scenarios, with "other" being given a one-in-five chance. However, individual responses varied a lot, and the self-reported confidence in these estimates was low: the median respondent rated their own confidence level as a 2, on a scale from 0 to 6. There was even enormous disagreement about the size of the threat: when asked about the size of existential risk from AI, respondents gave answers all the way from 0.1 percent to 95 percent.[5]

Much the same is true of issues around AI governance. In 2021 Luke Muehlhauser, a grantmaker in AI governance at Open Philanthropy, commented, "In the past few years, I've spent hundreds of hours discussing possible high-value intermediate goals with other 'veterans' of the longtermist AI governance space. Thus I can say with some confidence that there is very little consensus on which intermediate goals are net-positive to pursue."[6]

The second reason why we face such deep uncertainty is that, as well as weighing competing considerations we're aware of, we also need to try to take into account the considerations we haven't yet thought of. In 2002, when talking about the lack of evidence of Iraqi weapons of mass destruction, US Secretary of Defense Donald Rumsfeld declared, "There are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns—the ones we don't know we don't know."[7]

Rumsfeld's comment was lampooned as obscurantism at the time, and it even earned a Foot in Mouth Award, which the Plain English Campaign bestows each year for "a baffling comment by a public figure."[8] But he was actually making an important philosophical point: we should bear in mind there may be considerations that we aren't even aware of.

To illustrate, suppose that a highly educated person in the year 1500 tried to make the longterm future go well. They would be aware of some relevant

things, such as the persistence of laws, religions, and political institutions. But many issues wouldn't occur to them. The ideas that the earth's habitable life span could be a billion years and that the universe could be so utterly enormous, yet almost entirely uninhabited, would not have been on the table. Crucial conceptual tools for dealing with uncertainty, such as probability theory and expected value, had not yet been developed. They would not have been exposed to the arguments for a moral worldview in which the interests of all people are equal. They wouldn't have known what they didn't know.

The third reason why we face deep uncertainty is that, even in those cases where we know that a particular outcome is good to bring about, it can be very difficult to make that happen in a predictable way. Any particular action we take has a whole variety of consequences over time: some of these will be good, some will be bad, and many will be of unclear value. Nonetheless, ideally we should try to factor all the consequences we can into our decision.

When confronted with the empirical and evaluative complexity that faces us, it can be easy to feel clueless, as if there's nothing at all we can do. But that would be too pessimistic. Even if we're walking backwards into the future—and even if the terrain we're walking on is unexplored, it's dark and foggy, and we have few clues to guide us—nonetheless, some plans are smarter than others. We can employ three rules of thumb.

First, take actions that we can be comparatively confident are good. If we are exploring uncharted territory, we know that tinder and matches, a sharp knife, and first aid supplies will serve us well in a wide range of environments. Even if we have little idea what our expedition will involve, these things will be helpful.

Second, try to increase the number of options open to us. On an expedition, we would want to avoid getting stuck down a ravine we can't get out of, and if we weren't certain about the location of our destination, we would want to choose routes that leave open a larger number of possible paths. Third, try to learn more. Our expedition group could climb a hill in order to get a better view of the terrain or scout out different routes ahead.

These three lessons—take robustly good actions, build up options, and learn more—can help guide us in our attempts to positively influence the

long term. First, some actions make the longterm future go better across a wide range of possible scenarios. For example, promoting innovation in clean technology helps keep fossil fuels in the ground, giving us a better chance of recovery after civilisational collapse; it lessens the impact of climate change; it furthers technological progress, reducing the risk of stagnation; and it has major near-term benefits too, reducing the enormous death toll from fossil fuel–based air pollution.

Second, some paths give us many more options than others. This is true on an individual level, where some career paths encourage much more flexible skills and credentials than others. Though I've been very lucky in my career, in general, a PhD in economics or statistics leaves open many more opportunities than a philosophy PhD. As I suggested in Chapter 4, keeping options open is important on a societal level, too. Maintaining a diversity of cultures and political systems leaves open more potential trajectories for civilisation; the same is true, to an even greater degree, for ensuring that civilisation doesn't end altogether.

Third, we can learn more. As individuals, we can develop a better understanding of the different causes that I've discussed in this book and build up knowledge about relevant aspects of the world. Currently there are few attempts to make predictions about political, technological, economic, and social matters more than a decade in advance, and almost no attempts look more than a hundred years ahead. As a civilisation, we can invest resources into doing better—building mirrors that enable us to see, however dimly, into the future that lies behind us.

Keeping these high-level lessons in mind, let's talk about what to do, starting with the question of which priorities to focus on.

## Which Priorities Should You Focus On?

If you're on an expedition, there might be many problems facing you all at once: the tents leak; morale is low; a leopard is stalking you. You'd need to prioritise. The leaky tents might be annoying, but they're not as important as that leopard.

Similarly, when thinking about how to improve the world, the first step is to decide which problem to work on. When people are deciding how to do good, they often focus on a problem that is close to their heart, perhaps

because someone they know is affected by it. Others focus on problems that are especially salient. But if our aim is to do as much good as possible, these intuitions may be a poor guide, because the highest-impact actions may be much more effective than typical actions.

To get a sense for which kind of things we're choosing between, let's first take stock of the threats I've mentioned in the previous chapters. First, the lock-in of bad values, perhaps precipitated by artificial general intelligence or the dominance of a single world ideology. Second, the end of civilisation, which could be brought about by war involving nuclear weapons or bioweapons, or made more likely by technological stagnation, depleting fossil fuel reserves, or greatly warming the planet. What can we do in each of these areas?

For some issues, we can take somewhat robustly good actions. This is true for climate change and fossil fuel depletion, where we can draw on huge amounts of relevant research on their physical basis, their socioeconomic effects, and policies for mitigation and adaptation. And, crucially, we have a yardstick we can use to compare different interventions. We know we are winning against climate change if carbon dioxide emissions decline, and the more the better. Each of us can encourage clean-tech innovation through political advocacy or by funding or working for effective nonprofits like Clean Air Task Force and TerraPraxis.

Biosecurity and pandemic preparedness is another area where we can do things that are robustly good, like promoting innovation to produce cheap and fast universal diagnostics and extremely reliable personal protective equipment. Organisations like the Johns Hopkins Center for Health Security and the Bipartisan Commission on Biodefense are helping to promote pandemic preparedness solutions internationally.

General disaster preparedness also seems robustly good. This can include things like increasing food stockpiles; building bunkers to protect more people from worst-case catastrophes; developing forms of food production not dependent on sunlight in case of nuclear winter; building seed vaults with heirloom seeds that could be used to restart agriculture;[9] and building information vaults with instructions for creating the technologies necessary to rebuild civilisation.

In other areas, the key priorities are to build up options and learn more. This is true of many issues around AI. We do not yet know what the AGI

systems we're worried about are going to look like, except in their broad contours. This makes it hard to work on well-targeted solutions now, and because of the complex strategic situation, many well-intentioned attempts might even backfire.

The history of efforts to reduce AGI risk does illustrate, however, that there is at least one thing we can do in such a situation: building a field of morally motivated actors who can start reducing our uncertainty about what to do. Ten years ago, almost no one was working to positively steer the trajectory of AI. But there are now at least a hundred people working on this problem, and tens of millions of dollars are now spent on it every year.[10] Groups like the Center for Human-Compatible Artificial Intelligence and the Future of Humanity Institute have helped to build a field of researchers who are focused on safe AI development. The issue is also increasingly being taken seriously in technology policy, for instance by the Center for Security and Emerging Technology at Georgetown University in Washington, DC. This effort is still far too small, but it's growing.

The risk of great-power war is another example where field building and further research are key priorities. While there is a large body of work on the causes of war, we still have a lot to learn about practical ways to reduce the risks of war. For instance, we know that countries are more likely to go to war with each other if they have a long-standing rivalry or are geographic neighbours—especially if they have territorial disputes. But redrawing borders is hardly a feasible intervention, nor can we travel back in time to prevent countries from becoming rivals. And while we also know that democracies are less likely to fight each other, promoting democracy around the world is a major challenge. Given these uncertainties, identifying and training talented researchers and effective organizations who can improve our knowledge in this area strikes me as critical. Organisations like the Stockholm International Peace Research Institute may help us find the policies and programmes which, if implemented, give us the best chance at maintaining peace between great powers in the coming decades.

As well as improving our knowledge about particular issues, we can also try to get a better understanding of the implications of longtermism as a whole. For example, you can help find new crucial considerations. Perhaps there is an overlooked technology on the horizon that poses a grave threat to

the survival of civilisation. Perhaps some changes to the world's institutions and cultures would be valuable trajectory changes. Either of these would be enormously important to identify. These and other crucial issues are worked on at places like the Global Priorities Institute, the Future of Humanity Institute, and Open Philanthropy.[11]

How should we choose which of these problems are most pressing? In Chapter 2, I suggested using the significance, persistence, and contingency framework to measure a problem's importance.

But we should not *only* consider a problem's importance: some problems might be very important even though there is very little that we can do about it. We can break this down into two components. First, *tractability*. How many resources would it take to solve a given fraction of the problem? Some problems are intrinsically easier to make progress on than others. For example, the use of chlorofluorocarbons (CFCs) posed an enormous problem to the world by depleting the ozone layer.[12] But the problem turned out to be comparatively easy to solve: there were a small number of companies that needed to get on board and good substitutes for CFCs.[13] It was fifteen years between scientists first discovering that CFCs could deplete the ozone layer and the Montreal Protocol, which phased out chlorofluorocarbons and essentially ended the problem.[14]

For climate change, the difficulty of international cooperation and the lack of good substitutes for fossil fuels make the problem much harder.[15] But at least the nature of the problem—burning fossil fuels releases carbon dioxide—is very clear. This means we can create metrics by which we can more easily track progress on the problem. For other areas, like moral progress or the safe development of artificial intelligence, things are murkier. The nature of the problem is disputed, and there aren't such clear metrics by which we can track success.

The second component is *neglectedness*. The greater the number of people working on a problem, the more likely it is that the low-hanging fruit— the best opportunities to do good—will be taken. If you work on more neglected problems, you can make a bigger difference.

For instance, philanthropists now spend billions of dollars on climate advocacy every year, governments and companies spend hundreds of billions addressing climate change, and it is one of the problems of choice for most

young socially motivated people.[16] As I mentioned in Chapter 6, this is the main reason that the tide has started to turn on climate change. In contrast, issues around AI development are radically more neglected—though I noted that interest in the area is growing, philanthropic funding still amounts to only a few tens of millions of dollars a year, and there are only a couple of hundred people working in the area. This means that, if you can help make progress, you as an individual have the ability to be transformative in a way that is much harder in areas that have attracted more attention.

## How to Act

Assuming that you have chosen the problem you think is most pressing, what do you do next? People often focus on personal behaviour or consumption decisions. The suggestion, implicit or explicit, is that if you care about animal welfare, the most important thing is to become vegetarian; if you care about climate change, the most important thing is to fly less and drive less; if you care about resource overuse, the most important thing is to recycle and stop using plastic bags.

By and large, I think that this emphasis, though understandable, is a major strategic blunder for those of us who want to make the world better. Often the focus on consumption decisions is accompanied by a failure to prioritise. Consider, for example, the recent wave of advocacy for reducing plastic. The total impact this has on the environment is tiny. You would have to reuse your plastic bag eight thousand times in order to cancel out the effect of one flight from London to New York.[17] And avoiding plastic has only a tiny effect on ocean plastic pollution. In rich countries with effective waste management, plastic waste very rarely ends up in the oceans. Almost all ocean plastic comes from fishing fleets and from poorer countries with less-effective waste management.[18]

Some personal consumption decisions have a much greater impact than reusing plastic bags. One that is close to my heart is vegetarianism. The first major autonomous moral decision I made was to become vegetarian, which I did at age eighteen, the day I left my parents' home. This was an important and meaningful decision to me, and I remain vegetarian to this day. But how impactful was it, compared to other things I could do? I did it in large part because of animal welfare, but let's just focus on its effect on climate

change. By going vegetarian, you avert around 0.8 tonnes of carbon dioxide equivalent every year (a metric that combines the effect of different greenhouse gases).[19] This is a big deal: it is about one-tenth of my total carbon footprint.[20] Over the course of eighty years, I would avert around sixty-four tonnes of carbon dioxide equivalent.

But it turns out that other things you can do are radically more impactful. Suppose that an American earning the median US income were to donate 10 percent of that income, which would be around $3,000, to the Clean Air Task Force, an extremely cost-effective organisation that promotes innovation in neglected clean-energy technologies. According to the best estimate I know of, this donation would reduce the world's carbon dioxide emissions by an expected three thousand tonnes per year.[21] This is far bigger than the effect of going vegetarian for your entire life. (Note that the funding situation in climate change is changing fast, so when you read this, the Clean Air Task Force may already be fully funded. Giving What We Can keeps an up-to-date list of the best charities in climate and other areas.)

There are good reasons to become and stay vegetarian or vegan: doing so helps you be a better advocate for climate change mitigation and animal welfare, more able to avoid charges of hypocrisy; and you might reasonably think that avoiding causing unnecessary suffering is part of living a morally respectable life. But if your aim is to fight climate change as much as possible, becoming vegetarian or vegan is only a small part of the picture.

Emphasising personal consumption decisions over more systemic changes is often a convenient move for corporations. In 2019 Shell's chief executive, Ben van Beurden, gave a lecture in which he instructed people to eat seasonally and recycle more, lambasting people who eat strawberries in winter.[22] In reality, in order to solve climate change, what we actually need is for companies like Shell to go out of business. By donating to effective nonprofits, we can all make this kind of far-reaching political change much more likely.

Donations are more impactful than changing personal consumption decisions in other areas too. For example, in *Doing Good Better*, I argued that donating to the best global poverty charities is much more impactful than buying fair trade products. These examples are not a fluke. We should expect this pattern in almost all areas. The most powerful and yet simple reason is this: our consumption is not optimized for doing harm, and so by making

different consumption choices we can avoid at most the modest amount of harm we'd be otherwise causing; by contrast, when donating we can choose whichever action *best* reduces the harm we care about. We can have as big an impact as possible by taking advantage of levers such as affecting policy.

Moreover, for many of the problems I have discussed in this book, it is just not possible to make any difference by changing your consumption behaviour. While each of us can mitigate climate change through our everyday actions, this is not true for the risk of a great-power war, engineered pandemics, or the development of AI. However, we can all work on these problems by donating to effective nonprofits. Whatever else you do in life, donations are one way to do an enormous amount of good.

Beyond donations, three other personal decisions seem particularly high impact to me: political activism, spreading good ideas, and having children.

The simplest form of political activism is voting. On the face of it, it is improbable that voting could really do a lot of good. Every election I have ever voted in would have turned out the same whether I had voted or not, and that is almost certainly true for everyone reading this book. What this line of reasoning neglects is that, even if the chance that you influence an election is small, the *expected* value can still be very high.[23] If you live in the United States in a competitive state, the chance that your vote will flip a national election falls between one in one million and one in ten million. As a rule of thumb, governments typically control around a third of a country's GDP. In the United States, the federal government spends $17.5 trillion every four years. The spending priorities of administrations overlap substantially, so your vote may influence perhaps only 10 percent of the budget. Even so, multiply the small probability of your vote making a difference in a national election with the enormous impact if your vote *does* make a difference, and your vote in a competitive state would influence an expected $175,000. And this is just considering the money you might affect. A bigger effect could come from harder-to-quantify factors such as the likelihood that different candidates will start a nuclear war. So even though the probability of flipping an election is small, the payoff can be big enough to make voting worthwhile.

There are several caveats to this. First, many voters do not live in competitive states. If you live in a state that's certain to go to a particular candidate,

the expected value of voting might be tiny because the chance of your having an effect is so small. Second, to make your vote worthwhile, you need to do more than just turn up and vote; you need to be better informed and less biased than the median voter—otherwise you risk doing harm.

Many of the same arguments apply to other forms of political activism. Although the chance that you personally will make a difference by getting involved in a political campaign is small, the expected returns can be very high because if your campaign succeeds, the payoff could be very large.

Another way to improve the world is to talk to your friends and family about important ideas, like better values or issues around war, pandemics, or AI. This doesn't mean that you should promote these ideas aggressively or in a way that might alienate those you love. But discussion between friends has been shown to be one of the most effective ways to increase political participation,[24] and it is also probably a good way to get people motivated to work on some of the major problems of our time.

The final high-impact decision you can make is to consider having children. As I argued in Chapter 8, one mistake people sometimes make is to overemphasise the negative effects of having children and not to consider the benefits at all, both to the children and to the world. Although your offspring will produce carbon emissions, they will also do lots of good things, such as contributing to society, innovating, and advocating for political change. I think the risk of technological stagnation alone suffices to make the net longterm effect of having more children positive. On top of that, if you bring them up well, then they can be change makers who help create a better future. Ultimately, having children is a deeply personal decision that I won't be able to do full justice to here—but among the many considerations that may play a role, I think that an impartial concern for our future counts in favour, not against.

## Career Choice

So far, I have looked at ways that you can use your time and money to improve the long term. But by far the most important decision you will make, in terms of your lifetime impact, is your choice of career. Especially among young people, it has become increasingly common to strive for positive impact as a core part of one's professional life rather than as a sideshow. More

and more people don't just want money to pay their bills; they also want a sense of purpose and meaning.

This is why, as a graduate student, I cofounded 80,000 Hours with Benjamin Todd. We chose the name 80,000 Hours because that is roughly how many hours you have in your career: forty hours per week, fifty weeks per year, for forty years. Yet the amount of time that people normally spend thinking about their career is tiny in comparison. When that's combined with how poor existing career advice is, we end up with the outcome that a large proportion of people land in careers that are neither as fulfilling nor as impactful as they could be.

How, then, should you decide on a career? Again, we can return to our expedition metaphor. The three key lessons we identified were to learn more, build options, and take robustly good actions. These mirror the considerations that longtermists face when choosing a career:

1. **Learn:** Find low-cost ways to learn about and try out promising longer-term paths, until you feel ready to bet on one for a few years.
2. **Build options:** Take a bet on a longer-term path that could go really well (seeking upsides), usually by building the career capital that will most accelerate you in it. But in case it doesn't work out, have a backup plan to cap your downsides.
3. **Do good:** Use the career capital you've built to support the most effective solutions to the most pressing problems.

In reality, you'll be pursuing all of these priorities throughout your career, but each one will get different emphasis at different stages. Learning will tend to be most valuable early in your career. Building your options by investing in yourself and accruing career capital is most valuable in the early to middle stages of your career. Making a bet on how to do good is most valuable in the mid to late stages of your career. But your emphasis might move back and forth over time. For instance, a forty-year-old who decides to make a dramatic career change might go back into learning mode for a few years. And you might be lucky enough to find yourself with opportunities to have an enormous positive impact right out of college; if so, this framework shouldn't discourage you from doing that.

Let's first look at *learning*. People often feel a lot of pressure to figure out their best path right away. But this isn't possible. It's hard to predict where you'll have the best fit, especially over the long term, and if you're just starting out, you know very little about what jobs are like and what your strengths are. Moreover, even if you could find the best path now, it might change over time. The problems that are most pressing now could become less pressing in the future if they receive more attention, and new issues could be discovered. Likewise, you might find new opportunities to make progress that you hadn't anticipated.

Even your personal preferences are likely to change—probably more than you expect. Ask yourself, How much do you think your personality, values, and preferences will change over the next decade? Now ask, How much did they change over the previous decade? Intuitively, I thought they wouldn't change much over the next decade, but at the same time I think they changed a lot over the previous decade, which seems inconsistent. Surveys find similar results, which suggests that people tend to underestimate just how much they will change in the future.[25]

All of this means that it's valuable to view your career like an experiment—to imagine you are a scientist testing a hypothesis about how you can do the most good. In practical terms, you might follow these steps:

1. Research your options.
2. Make your best guess about the best longer-term path for you.
3. Try it for a couple of years.
4. Update your best guess.
5. Repeat.

Rather than feeling locked in to one career path, you would see it is an iterative process in which you figure out the role that is best for you and best for the world. The value of treating your career like an experiment can be really high: if you find a career that's twice as impactful as your current best guess, it would be worth spending up to half of your entire career searching for that path. Over time, it will become clearer whether you have found the right path for you. For many people, I think it would be reasonable to spend

5 percent to 15 percent of their career learning and exploring their options, which works out to two to six years.

Kelsey Piper provides one example of the value of learning early about your options. In order to test out her potential as a writer, while in college she wrote one thousand words a day for her blog.[26] It turned out that she was good at it. Blogging helped her figure out that writing was the right path for her and helped her to eventually get a job at Vox's Future Perfect, which covers topics relevant to effective altruism, including global poverty, animal welfare, and the longterm future.

When you are thinking about exploration, I think it is good to aim high, to focus on "upside options"—career outcomes that have perhaps only a one-in-ten chance of occurring but would be great if they did. Shooting for the moon is not always good advice. However, if you want to have a positive impact on the world, there's a strong case to be made for aiming high. Even if there is a small chance of success, the expected value of focusing on upside options can be great, and, crucially, there is a large skew in outcomes. In many fields, the most successful people are responsible for a large fraction of the impact; for example, various studies have found that the top 20 percent of contributors produce a third to a half of the total output.[27]

Even though focusing on upside options when you are exploring is very valuable, you should also limit the risk that you could do harm. Because we are so uncertain about longterm effects, there is an increased risk of doing harm, so you should take this consideration seriously. In a slogan: target upsides but limit downsides.

The next thing to consider on your career path is *building options* by investing in yourself. In a lot of fields, people's productivity peaks between ages forty and fifty.[28] So investing in career capital, in the skills and networks you need to have a big impact, is a top priority early in your career. Some of the skills you could focus on include the following:[29]

- Running organisations
- Using political and bureaucratic influence to change the priorities of an organisation
- Doing conceptual and empirical research on core longtermist topics

- Communicating (for example, you might be a great writer or podcast host)
- Building new projects from scratch
- Building community; bringing together people with different interests and goals

Investing in yourself can pay off in unanticipated ways. For example, based on 80,000 Hours's advice, Sophie decided not to apply to medical school and instead shifted her focus to global pandemics. She found funding for a master's degree in epidemiology to build career capital in the area. When COVID-19 broke out, she found a neglected solution: challenge trials, which can greatly speed up the development of vaccines by deliberately infecting healthy and willing volunteers with the novel coronavirus in order to test vaccine efficacy. So she co-founded 1DaySooner, a nonprofit that signed up thousands of volunteers for human challenge trials in order to speed up vaccine approval. The world's first challenge trial for COVID vaccines started in the UK in early 2021.[30]

There is sometimes a trade-off between exploring and investing. This is particularly clear in academia. If I wanted to try out a different job and quit academic philosophy for a few years, that would probably be the end of my philosophy career—in my field, once you leave there is no way back. But things are not usually as clear-cut as this, and building career capital does not always preclude exploring later on.

The final consideration for choosing a career is the one we ultimately care about: *doing good*. For most people, the opportunity to have a lot of impact comes later in their career, once they have gained career capital. But sometimes you might come across a great opportunity to do good right away. For instance, Kuhan Jeyapragasan realised that his position as a student at Stanford University gave him a great platform for spreading awareness of important ideas. He helped to start the Stanford Existential Risk Initiative, which has helped hundreds of people learn about risks to humanity's longterm future.

In large part, how much good you do depends on the problem you choose to work on. As I argued earlier, there are probably very large differences in impact between problem areas, so making this choice carefully is crucial. The immediate impact you have will also be determined by the quality of the project you are working on, your seniority, and the strength of your team.

The "learn more, build options, do good" framework is generally useful for anyone deciding what to do with their career. But the specific path that works best for *you* depends on your *personal fit*. Some people are happiest locked away for months on end researching abstruse topics in economics or computer science, while others excel at managing a team or communicating ideas in a simple and engaging way.

You might also have some unique opportunities that other people don't have. Marcus Daniell is a professional tennis player from New Zealand. He is one of the top fifty doubles players in the world, and he won a bronze medal in doubles at the 2021 Tokyo Olympics. After learning about effective altruism, Marcus set up High Impact Athletes, which encourages professional athletes to donate to effective charities working on global development, animal welfare, and climate change. People who have donated through High Impact Athletes include Stefanos Tsitsipas, the current number four tennis player in the world, and Joseph Parker, a former world heavyweight champion boxer and sparring partner for Tyson Fury. The opportunity to set up High Impact Athletes was unique to Marcus; his network allowed him to try out something new and set up an organisation with lots of potential upside.

Isabelle Boemeke's story is in some ways similar. She 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 because of its unpopularity, she pivoted to using her social media following to advocate for nuclear power. Of course, I'm not recommending professional tennis or fashion modelling as reliably high-impact careers, but these examples illustrate the importance of focusing on where you personally, with all your unique skills and abilities, can make the biggest difference on the world's most pressing problems. It would, for instance, have made little sense for Marcus or Isabelle to retrain as an epidemiologist or a climate scientist.

For many people, personal fit can mean the best way of contributing is through donations: you work in a career you love and excel at, and even if the work itself is not hugely impactful, you can make an enormous difference with your giving. This was true of John Yan. After learning about

effective altruism and thinking about his career options, he decided to continue as a software engineer and donate a significant fraction of his income to effective charities as a member of Giving What We Can.[31]

Personal fit is a crucial determinant of your career's impact—it is a force multiplier on the direct impact you have and on the career capital that you gain. As mentioned before, outcomes are heavily skewed. If you can be in the top 10 percent of performers in a role rather than in the top 50 percent, this could have a disproportionate effect on your output. Being particularly successful in a role also gives you more connections, credentials, and credibility, increasing your career capital and leverage.

Personal fit is, in addition, one of the main ingredients of job satisfaction. People often associate altruism with self-sacrifice, but I think that for the most part, that is the wrong way to think about it. For me personally, since I started trying to do the most good with my life, I feel that my life is more meaningful, authentic, and autonomous. I am part of a growing community of people trying to make the world a better place, and many of these people are now among my closest friends. Effective altruism has added to my life, not subtracted from it. There is, moreover, a pragmatic reason to do a job you enjoy: it makes your impact sustainable over the long term. You want to be able to sustain your commitment to doing good for over forty years rather than think about how you can do as much good as possible this year. The risk of burnout is real, and you will work better with other people and be more productive if you are not stressed or depressed.

## Doing Good Collectively

I've argued that positively influencing the longterm future is a key moral priority of our time. But it's not the only thing that matters. We should try to make the longterm future better in the context of living a rounded ethical life.

As part of this, it's particularly important to avoid doing harm. History is littered with people doing bad things while believing they were doing good, and we should do our utmost to avoid being one of them. For example, consider the Animal Rights Militia, which in the 1980s and '90s in the UK sent letter bombs to members of Parliament, including the prime minister at the time, and used bombs to set fire to buildings across the UK. Those behind these actions presumably thought they were acting morally—doing what

was needed to reduce the suffering of animals. But they were wrong, and not just in this instance: doing significant harm to serve the greater good is very rarely justified. Here is why.

First, naive calculations that justify some harmful action because it has good consequences are, in practice, almost never correct. The Animal Rights Militia might have thought they were doing what was best for animals, but in reality they were hindering the cause by tainting it with violent extremism. This is particularly true when we consider that there are often a wide variety of ways of achieving a goal, many of which do not involve doing harm. The best alternative for the Animal Rights Militia wasn't sitting at home and doing nothing: it was engaging in peaceful and nonviolent protest and campaigning.

Second, plausibly it's wrong to do harm even when doing so will bring about the best outcome. This is an issue that divides what are called "consequentialists" and "nonconsequentialists" in moral philosophy. Even if you are sympathetic to consequentialism—in which the ends are all that ultimately matter—given the difficulty of ethics, you should not be certain in that view. And when we are morally uncertain, we should act in a way that serves as a best compromise between different moral views.[32] If one reasonable view says that avoiding harm is very important, then we should put significant weight on that when we act.

Similar considerations apply to other commonsense moral considerations. You might reason in a particular case that lying would produce the best consequences, but lying has many indirect negative effects that are difficult to observe, and it's plausibly intrinsically wrong too. So, in practice, I think it makes sense to almost never lie, even when it seems like doing so would be for the best. For similar reasons, one should strive to be a good friend and family member and citizen, to act kindly, and to cultivate a habit of cooperation—even if, in any given situation, it is not clear why this would lead to the best possible outcome. In these ways, I see longtermism as a supplement to commonsense morality, not a replacement for it.

A different way in which naive expected-value reasoning can lead us astray is if we think too individualistically, paying attention only to what we as individuals can achieve rather than thinking in terms of what the whole community of people engaged in longtermism can do.

I have seen the importance of group action firsthand through the effective altruism community. Since it was formed a decade ago, this community has grown to thousands of members who share information and opportunities, have their own online forum to discuss the latest ideas, and provide friendship and social support for one another. Undoubtedly, the community is more than the sum of its parts: we can achieve far more by working together than we would if we each tried to do good on our own. Importantly, because this community has a shared aim of doing the most good, I have reasons to help others in the community even if I do not receive anything in return.

The fact that we each act as part of a wider community warrants a "portfolio approach" to doing good—taking the perspective of how the community as a whole can maximize its impact. Then you can ask what you can do to move the community closer to an ideal allocation of resources, given everyone's personal fit and comparative advantage. Taking a community perspective, the primary question becomes not "How can I personally have the biggest impact?" but "Who in the community is relatively best placed to do what?" For example, my colleague Greg Lewis believes that AI risk is the most important issue of our time. But he thinks the risk from engineered pandemics is also important, and because he has a medical degree, it makes more sense for him to focus on that threat and let others focus on AI.

The portfolio approach can also give greater value to experimentation and learning. If one person pursues an unexplored path to impact (such as an unusual career choice), everyone else in the community gets to learn whether that path was successful or not. It can also give much greater value to specialisation: a community of three people might need only generalists, but a community of thousands will need people with particular specialist skills.

The portfolio approach also makes it easier to see how you can have an impact. If you only consider what you personally might be able to achieve, it is easy to feel powerless in the face of huge international problems like climate change and engineered pathogens. But if you instead ask "Would we make progress on the threat from engineered pandemics if there were hundreds of motivated and smart people working on it?" I think it becomes clear that the answer is yes.

## Building a Movement

This chapter has discussed many ways you can directly have impact. But you can also go "meta": spread the idea of longtermism itself and convince others to care about future generations, to take the scale of the future seriously, and to act to positively influence the long term. You can do this by writing, organizing, talking to people you know, or getting involved with organisations such as 80,000 Hours and the Centre for Effective Altruism, where movement building is a component of their work.

Spreading these ideas can be an enormously powerful way of having an impact. Suppose that you convince just one other person to do as much good as you otherwise would have done in your life. Well, then you've done your life's work. Convince two other people, and you've tripled your impact.

Of course, we can take this reasoning too far. There are limits to how big a longtermist movement could be. And ultimately, movement building isn't enough: we need to actually solve the problems I've discussed.

But the nascency of longtermism suggests that developing and spreading ideas around it should be a core part of the movement's portfolio. For many previous social movements, change took time. The first public denouncement of slavery by the Quakers—the Germantown petition—was in 1688.[33] The Slavery Abolition Act in the British Empire was passed only in 1833, and several countries abolished slavery after 1960. Success took hundreds of years.

So, too, with feminism. Mary Wollstonecraft is often regarded as the first English-language feminist.[34] Her seminal work, *A Vindication of the Rights of Woman*, was published in 1792. The United States and the UK only gave men and women equal voting rights in 1920 and 1928, respectively, and it was only in 1971 that Switzerland did the same.[35] And of course, there is still much further to go on women's rights.

We may not see longtermism's biggest impacts in our lifetimes. But by advocating for longtermism, we can pass the baton to those who will succeed us—those who might run faster, see farther, and achieve more than we ever could. They will have the benefits of decades' more thought on these issues. And perhaps crucial moments of plasticity, when the direction of civilisation will be set, will occur during their lives rather than ours.

Recent history should give us hope that the world will start taking the interests of future generations seriously. Environmentalists have made the wellbeing of future generations salient in a way that has had real impact. To take just one example: After decades of campaigning, in 1998 the Greens became part of the coalition government in Germany, and in 2000, they introduced landmark legislation that would almost singlehandedly underwrite the global solar industry's growth, making Germany the world's largest solar market. By 2010, Germany accounted for nearly half of the global market for solar deployment.[36] From the perspective of providing power to Germany alone—a northern-latitude and fairly cloudy country—this made little sense. But from a global perspective, it was transformative. Thanks to this and other subsidy schemes introduced around the same time, the cost of solar panels fell by 92 percent between 2000 and 2020.[37] The solar revolution that we're about to see is thanks in large part to German environmental activism.[38]

I've seen successes from those motivated explicitly by longtermist reasoning, too. I've seen the idea of "AI safety"—ensuring that AI does not result in catastrophe even after AI systems far surpass us in the ability to plan, reason, and act (see Chapter 4)—go from the fringiest of fringe concerns to a respectable area of research within machine learning. I've read the UN secretary-general's 2021 report, *Our Common Agenda*, which, informed by researchers at longtermist organisations, calls for "solidarity between peoples and future generations."[39] Because of 80,000 Hours, I've seen thousands of people around the world shift their careers towards paths they believe will do more longterm good.

But we should not be complacent. There are enormous challenges ahead. We need to decarbonise the economy over the next fifty years, even as energy demand triples.[40] We need to reduce the risks of war between great powers, of the use of engineered pathogens, and of AI-assisted perpetual global totalitarianism. And at the same time, we need to ensure that the engine of technological progress keeps running.

If we are to meet these challenges and ensure that civilisation at the end of this century is pointed in a positive direction, then a movement of morally motivated people, concerned about the whole scope of the future, is a necessity, not an optional extra.

Who should this movement consist of? Well—if not you, then who?[41]

Positive moral change is not inevitable. It's the result of long, hard work by generations of thinkers and activists. And if there's any change that's not inevitable, it's concern for future people—people who, by virtue of their location in time, are utterly disenfranchised in the world today.

If we are careful and far-sighted, we have the power to help build a better future for our great-grandchildren, and their great-grandchildren—down through hundreds of generations. But we cannot take such a future for granted. There's no inevitable arc of progress. No *deus ex machina* will prevent civilisation from stumbling into dystopia or oblivion. It's on us. And we are not destined to succeed.

Yet success is possible—at least if people like you rise to the challenge. You may have more power than you realise. If your income is more than $20,000 per year (post-tax, with no dependents), then you are in the richest 5 percent of the world's population, even after adjusting for the fact that money goes further in lower-income countries.[42] And you probably live in one of the more powerful countries in the world, where you can campaign to change the attitudes of your conationals and the policies of your government.

If you've read this far, then probably you *care*, too. The last ten chapters have not been easy. Since you've made it through discussions of impossibility theorems in population ethics and of weighing chicken suffering against human happiness, you probably were convinced enough by my arguments in the first chapters that you wanted to know how it would all pan out—what the practical upshot would be. If there's ever anyone who will take action on behalf of future generations, it's you.

But can one person make a difference? Yes. Mountains erode because of individual raindrops. Hurricanes are just the collective movement of many tiny atoms. Abolitionism, feminism, and environmentalism were all "merely" the aggregate of individual actions. The same will be true for longtermism.

We've met some people who made a difference in this book: abolitionists, feminists, and environmentalists; writers, politicians, and scientists. Looking back on them as figures from "history," they can seem different from you and me. But they weren't different: they were everyday people, with their own problems and limitations, who nevertheless decided to try to shape the history they were a part of, and who sometimes succeeded. You can do this, too.

Because if not you, who? And if not now, when?

Out of the hundreds of thousands of years in humanity's past and the potentially billions of years in her future, we find ourselves living *now*, at a time of extraordinary change. A time marked by the shadow of Hiroshima and Nagasaki, with thousands of nuclear warheads standing ready to fire. A time when we are burning through our finite fossil fuel reserves, producing pollution that might last hundreds of thousands of years. A time when we can see catastrophes on the horizon—from engineered pathogens to value lock-in to technological stagnation—and can act to prevent them.

This is a time when we can be pivotal in steering the future onto a better trajectory. There's no better time for a movement that will stand up, not just for our generation or even our children's generation, but for all those who are yet to come.

# Acknowledgements