In the future, if we spread to the stars, we will again be separated. The galaxy is like an archipelago, vast expanses of emptiness dotted with tiny pinpricks of warmth. If the Milky Way were the size of Earth, our solar system would be ten centimetres across and hundreds of metres would separate us from our neighbours. Between one end of the galaxy and the other, the fastest possible communication would take a hundred thousand years; even between us and our closest neighbour, there-and-back communication would take almost nine years.[60]

In fact, if humanity spreads far enough and survives long enough, it will eventually become impossible for one part of civilisation to communicate with another. The universe is composed of millions of groups of galaxies.[61] Our own is called, simply, the Local Group. The galaxies within each group are close enough to each other that gravity binds them together forever.[62] But, because the universe is expanding, the groups of galaxies will eventually be torn apart from each other. Over 150 billion years in the future, not even light will be able to travel from one group to another.[63]

The fact that our time is so unusual gives us an outsized opportunity to make a difference. Few people who ever live will have as much power to positively influence the future as we do. Such rapid technological, social, and environmental change means that we have more opportunity to affect when and how the most important of these changes occur, including by managing technologies that could lock in bad values or imperil our survival. Civilisation's current unification means that small groups have the power to influence the whole of it. New ideas are not confined to a single continent, and they can spread around the world in minutes rather than centuries.

The fact that these changes are so recent means, moreover, that we are out of equilibrium: society has not yet settled down into a stable state, and we are able to influence *which* stable state we end up in. Imagine a giant ball rolling rapidly over a rugged landscape. Over time it will lose momentum and slow, settling at the bottom of some valley or chasm. Civilisation is like this ball: while still in motion, a small push can affect in which direction we roll and where we come to rest.

CHAPTER 2

# You Can Shape the Course of History

## Prehistory's Impact on Today

Human beings have been making choices with longterm consequences for tens of thousands of years. Consider: Why is Africa home to so many more species of megafauna—large animals like elephants and giraffes—than the rest of the world?[1] You might think, as I did before learning about this topic, that the answer has to do with Africa's particular environment. But that's not right. Fifty thousand years ago, a great variety of megafauna roamed the planet.

Consider the glyptodonts, a group of armadillo-like herbivores that lived in South America for tens of millions of years.[2] The largest glyptodonts were as big and heavy as cars.[3] Their bodies were encased in a giant shell, they had a bone helmet, and some of them had club-shaped tails adorned with spikes.[4] They looked like giant capybaras dressed up as armoured trucks. They went extinct around 12,000 years ago.[5]

Or consider megatherium, a giant ground sloth and one of the largest land mammals to have ever lived, rivalling the Asian elephant in size.[6] It went extinct 12,500 years ago.[7] Or *Notiomastodon*, a genus of elephant-like animals with giant tusks that evolved two million years ago and went extinct 10,000 years ago.[8] Or the dire wolf, the largest known canine to have lived, which, having lost its giant herbivorous prey, went extinct 13,000 years ago.[9] All these species lived in South America, along with dozens of other megafauna species that are no longer with us.

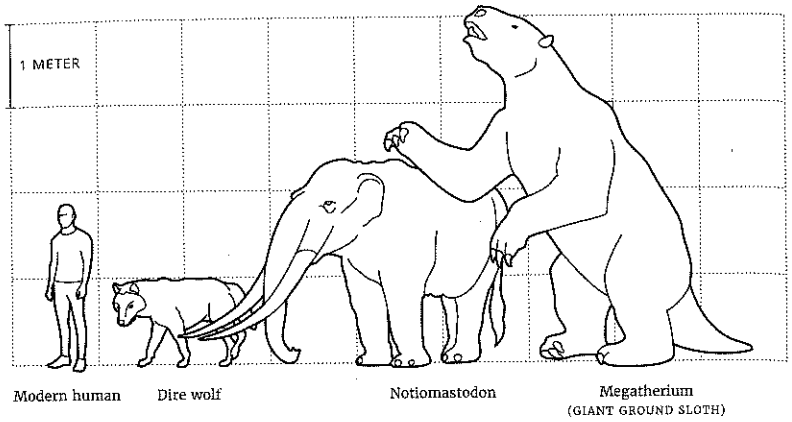


*Figure 2.1. Some specimens of now-extinct megafauna drawn at scale in comparison with a modern human.*

There is a heated debate over what caused the extinctions of megafauna. Some scientists believe that natural climate change was the main driver, some believe that humans were the culprit, and some believe it was a mix of humans and climate change.[10] In my view, the evidence is clear that humans often played a decisive role: most of these megafauna survived over a dozen similarly sized climatic changes in the past;[11] smaller animals did not go extinct at nearly the same rate as megafauna;[12] and the timing of their extinction usually coincides with humans' arrival into their habitats.[13] Though perhaps helped by climate change, it was hunting and the disruption of natural environments caused by human activity that killed them off. Unlike megafauna on other continents, African megafauna evolved alongside humans and so were better prepared for *Homo sapiens* as a predator.

The extinction of these megafauna was probably an irrevocable change to the world, made by humans with extremely primitive technology. It meant we lost, for all time, many beautiful and unique species. And *Homo sapiens* are not only implicated in the extinction of giant sloths and canines: we are also the prime suspect in the end of our human cousins, the Denisovans and the Neanderthals, who likely died out as a result of both competition and interbreeding.[14] There is now only one *Homo* species on the planet, but there could have been many.



Early humans made other choices with longterm consequences, too. Early agriculturalists, for example, burned down vast swathes of forest to create plains for farming and paddies for rice irrigation.[15] This preindustrial deforestation had a lasting impact. Because carbon dioxide remains in the atmosphere for so long, the planet is, as a result of the actions of our ancestors, slightly warmer today.[16]

Just as actions taken by our ancestors thousands of years ago shaped the present day, so too will decisions we make today shape the future thousands of years hence. But to justify taking a longterm view of our decisions, what matters is not only whether we can impact the future but whether we can adequately foresee what those impacts will be. We don't need to predict every detail, nor could we if we tried. But if we want to make the future better, we need to identify actions that have positive effects on balance over very long timescales.

Our distant ancestors could not predict their longterm impact on the world. Hunter-gatherers did not know they were driving species to extinction. Early agriculturalists could not guess that deforestation would warm the planet, nor what the consequences of this warming would be.

But we in the modern era can do better. Clearly, there's still much we don't know, but in the last few centuries especially, we've learned a lot. If early agriculturalists had had our understanding of climate physics, they could have foreseen some of the geophysical impacts of burning forests; if hunter-gatherers had had our knowledge of ecology and evolutionary biology, they would have understood what it is for a species to go extinct and the potentially irrevocable loss that was at stake. With careful investigation and appropriate humility, we can now start to assess the effects of our actions over very long timescales.

In this chapter, I'll present a framework for assessing the longterm value of an event. The chapters that follow apply this framework to events that I think we, today, can foreseeably influence for the better.

## A Framework for Thinking About the Future

Consider some state of affairs that people could bring about, like the nonexistence of the glyptodonts. We can assess the longterm value of this new

state of affairs in terms of three factors: its significance, its persistence, and its contingency.[17]

*Significance* is the average value added by bringing about a certain state of affairs. How much worse is the world, at any one time, because the glyptodonts are extinct? In assessing this, we would want to attend to all relevant aspects of the glyptodonts' extinction: the intrinsic loss of a species on the planet, the loss to humans who could have used their shells or eaten their meat, and the impact on the ecosystems the glyptodonts inhabited.

The *persistence* of a state of affairs is how long that state of affairs lasts, once it has been brought about. The nonexistence of the glyptodonts may be exceptionally persistent, starting 12,000 years ago and lasting until the end of the universe.[18] It would only fail to be exceptionally persistent if, at a future time, we were to bring them back.

Technology may make this possible. There are current efforts to "de-extinct" certain species, like the woolly mammoth, by extracting DNA from their remains and editing that DNA into the cells of similar modern animals, like elephants.[19] However, even if successful, these efforts would not truly bring back the original creatures: instead, they would produce a hybrid—an animal that looks a lot like the extinct animal but is not genetically the same. Should future generations try to bring back the glyptodonts, they would probably face similar challenges.

The final aspect of the framework is *contingency*. This is the most subtle part of the framework. In English the word "contingency" has a few different meanings; in the sense I'm using it, an alternative term would be "noninevitability." Contingency represents the extent to which a state of affairs depends on a small number of specific actions. If something is very contingent, then that change would not have otherwise occurred for a very long time, or ever. The existence of the novel *Jane Eyre* is very contingent: if Charlotte Brontë had not written it, that precise novel would never have been written by someone else. Agriculture is less contingent because it emerged in multiple locations independently.

If something is very noncontingent, then the change would have happened soon anyway, even without the individual's action. Knowledge of calculus was not very contingent because Leibniz independently discovered it just a few years after Newton did. Considering contingency is crucial



because if you make a change to the world but it's a change that would have simply happened soon afterward anyway, then you have not made a longterm *difference* to the world.

Though it's hard to be confident, my guess is that the extinction of the glyptodonts was not very contingent. Even if the hunters who killed off the last of them had not done so, then probably some other group of hunters, at some later time, would have. In order to prevent the glyptodonts' extinction, those hunters would have had to promote a norm that the glyptodonts should be protected and this norm would have had to be passed down the generations, and adhered to, until the present day. This would not be impossible to pull off, but it does seem difficult.

Multiplying significance, persistence, and contingency together gives us the longterm value of bringing about some state of affairs. Because of this, we can make intuitive comparisons between different longterm effects on these dimensions. For example, between two alternatives, if one is ten times as persistent as the other, that will outweigh the other being eight times as significant. Because the potential scale of the longterm future is so great—millions, billions, or even trillions of years—our attention should be, first, on what states of affairs might be the most persistent. Then, afterwards, we can think about significance and contingency.

**Table 2.1. The Significance, Persistence, Contingency Framework**

| | |
|---|---|
| **Significance** | What's the average value added by bringing about a certain state of affairs? |
| **Persistence** | How long will this state of affairs last once it has been brought about? |
| **Contingency** | If not for the action under consideration, how briefly would the world have been in this state of affairs (if ever)? |

Note: For more details, see Appendix 3.

To see how this framework can be used to guide our decisions today, let's return to the metaphor of humanity as an imprudent teenager. Looking back at our own individual teenage years, what choices mattered most? Plausibly, it's those whose effects were the most persistent, affecting the whole course of our lives; most significant, making the biggest difference to our wellbeing

at any one time; and most contingent, causing an effect that would not have happened anyway at some later date.

Some choices I made as a teenager did not have persistent effects: my plans for the weekend made a difference to that weekend but usually didn't shape the course of my life. The effects of other choices were not that contingent. Like many teenagers, I cared about firsts—first drink, first time having sex. But ultimately, such firsts would have happened at some point regardless, and looking back, the precise timing did not matter much. Finally, some effects, though persistent and contingent, just weren't that significant. I chose not to get braces to close the gap between my two front teeth because at the time I believed that a gap brings good luck. I still have the gap today, but as far as I can tell, it has not significantly affected my life.

Other decisions I made mattered a lot. I was reckless as a teenager and sometimes went "buildering," also known as urban climbing. Once, coming down from the roof of a hotel in Glasgow, I put my foot on a skylight and fell through. I caught myself at waist height, but the broken glass punctured my side. Luckily, it missed all internal organs. A little deeper, though, and my guts would have popped out violently, and I could easily have died. I still have the scar: three inches long and almost half an inch thick, curved like an earthworm. Dying that evening would have prevented all the rest of my life. My choice to go buildering was therefore an enormously important (and enormously foolish) decision—one of the highest-stakes decisions I'll ever make.

More mundanely, I could easily have exposed myself to a different set of intellectual influences, which would have set me on a very different path in life. All my close friends studied medicine—the standard path for smart, socially minded teenagers in Scotland—and I considered it for myself. If I had not studied philosophy at school, and if I hadn't had such an engaged and passionate teacher, Jeremy Hall, I would probably not have studied it at university or pursued it as a career. I expect that a career in medicine would have been fulfilling, but it probably would not have exposed me to the moral arguments that led me to the path I've taken—a difference which, from my current perspective, would have been a major loss.

Looking back, it's clear that, for many of my teenage choices, what mattered most was not the fun I had at the time—whether buildering was a



thrill (it was) or whether studying medicine at Edinburgh involved better parties. Rather, what mattered most was the impact of these choices on the rest of my life, whether I was risking death or altering the values that would guide my future self.

The risk of death I bore as a teenager and the intellectual influences that shaped my life mirror the two main ways in which we can impact the longterm future. First, we can affect humanity's duration: ensuring that we survive the next few centuries affects how many future generations there are. That is, we can help *ensure civilisation's survival.* Just as my teenage decisions to gamble with my life were among the most consequential I've ever made, so too are our decisions about how to handle risks of extinction or unrecovered civilisational collapse among the most consequential decisions that we as a society make today.

Second, we can affect civilisation's average value, changing how well or badly life goes for future generations, potentially for as long as civilisation lasts. That is, we can *change trajectory,* trying to improve the quality of future people's lives over the life span of civilisation.[20] Just as the intellectual influences I was exposed to as a teenager shaped the whole rest of my life, so, too, I will argue, the values that humanity adopts in the next few centuries might shape the entire trajectory of the future.[21]

These two ideas structure the book. Part II of this book looks at trajectory changes, focusing in particular on changing society's values. Within this,

**TWO WAYS TO IMPROVE THE FUTURE**

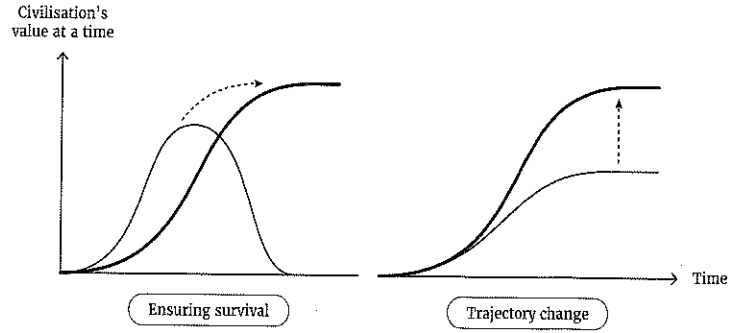


*Figure 2.2. We can make the future better in two ways: by averting permanent catastrophes, thereby ensuring civilisation's survival; or by changing civilisation's trajectory to make it better while it lasts.*

Chapter 3 argues for the significance and contingency of value changes, focusing on the abolition of slavery as a case study. Chapter 4 argues for the persistence of values, suggesting that new technology, in particular advanced artificial intelligence, could enable those in power to lock in their values indefinitely. Whether the future is governed by values that are authoritarian or egalitarian, benevolent or sadistic, exploratory or rigid, might well be determined by what happens this century.

Part III looks at three ways of ensuring survival, dedicating a chapter to each. The first way is to prevent direct risks of human extinction; I focus on engineered pandemics. The second is to prevent the unrecovered collapse of civilisation; I focus on risks from nuclear war and extreme climate change. The third is technological stagnation, which could increase the risks of both extinction and collapse. Along the way, I discuss the persistence and contingency of the end of civilisation.

The question of the *significance* of the end of civilisation raises philosophical issues. Broadly, ensuring survival increases the quantity of future life; trajectory changes increase its quality. But you might not care much about sheer quantity. If there's no longer anyone around to care, why should it matter if civilisation has ended? And maybe, on balance, the future is more bad than good. If these worries were correct, then the longtermist priority should be to increase the average value of future civilisation rather than its duration. Improving our trajectory would be more important than ensuring survival.

Part IV tackles these issues. I argue both that we should think of the nonexistence of future generations as a moral loss, if the people in them would have sufficiently good lives, and that we should expect the future to be more good than bad, on balance. Ensuring survival is therefore just as great a priority as improving our trajectory.

Part V turns to action. Longtermism is not just abstract philosophical speculation. It's an idea that people are putting into practice today. Chapter 10 looks at what some people are doing today to try to make the long term better, and how you can help.[22]

## Thinking in Bets

When thinking about the changes that we could make to the world, we will not know how long they will last or how significant or contingent they will



be. So we need a way of making decisions in the face of uncertainty. The most widely accepted account of how to do so is expected value theory.

Over the course of writing this book, I was repeatedly and viscerally reminded of the idea of expected value theory by my housemate at the time, Liv Boeree. Liv is one of the most successful female poker players of all time—a European Poker Tour and World Series champion. Her understanding and internalisation of the idea of expected value—or "EV," as she calls it—is critical to her success.

There are three aspects to expected value. First, probabilities. Rather than thinking that a three-of-a-kind poker hand is "very unlikely," Liv knows that the chance of getting one, before any cards are dealt, is about 5 percent; if the first two cards she's dealt are a pair, this probability rises to about 12 percent.[23] Though both probabilities are small, the difference between them can easily be enough to affect your decisions at the poker table.

What's striking about Liv is that she applies this same probabilistic thinking to other areas of her life, too. She and her partner, Igor (another poker player), will happily discuss the probability that they'll still be together after ten years. (It's currently at 80 percent.)

It can feel unnatural to apply probabilities to areas of life where chances aren't easily quantified. But it means we can have more nuanced and accurate views about the world. It's a way of thinking more precisely. "People often think something definitely will or definitely won't happen—as zero percent chance or a hundred percent chance," Liv told me. "But of course almost everything falls in between. Or else they use vague language like 'a fair chance.' But a 'fair chance' means very different things to different people."

She's right. One study found that people interpret the phrase "might happen" to refer to anything between 10 percent and 60 percent probability, and "a serious possibility" as all the way from 30 percent to 90 percent.[24] This vagueness can have momentous implications. In 1961, when President John F. Kennedy asked the military for advice on whether to invade Cuba at the Bay of Pigs, he was told that the plan had a "fair chance" of success. Quite reasonably, Kennedy took that to be a positive assessment. But the author of the words "fair chance" later said that he meant that there was only about a 30 percent chance of success.[25] The operation failed dramatically.

The second aspect of expected value is assigning values to outcomes. For professional poker players, this is comparatively easy: they can just look at their financial returns. But financial returns are not in general the right measure of value. If you need £1000 to pay for a life-saving operation, then the difference in value for you between getting nothing and getting £1000 is much greater than the difference in value between getting £1000 and getting £2000. The value that we assign to outcomes should be based on whatever it is we *ultimately* care about, such as people's wellbeing.

Precisely assigning value to different outcomes can be difficult, but we often only need very rough comparisons in order to make a decision. Suppose that there are two different drugs that could cure a patient's ailment, with different side effects. The first will certainly cause a mild headache; the second has a one-in-ten risk of causing a fatal heart attack. It's hard to know exactly how much worse death is than a mild headache. But, apart from exceptional cases, it's certainly more than ten times worse.

This brings us to the third aspect of expected value theory, which is measuring how good or bad a decision is by its expected value. This can be intuitive: in the two-drugs example I just gave, the first drug is the better choice; death is more than ten times as bad as a mild headache, so a 10 percent risk of death is sufficient to outweigh a guarantee of a headache. We can calculate the expected value of a decision as follows. First, we list each possible outcome of the decision. Next, we assign a probability and a value to each outcome, which we then multiply together. Finally, we add up all the probability-times-value products.

Liv and Igor make bets against each other all the time, and they decide whether to take them on the basis of expected value. To take one real-life example, suppose that Liv and Igor are at a pub, and Liv bets Igor that he can't flip and catch six coasters at once with one hand. If he succeeds, she'll give him £3; if he fails, he has to give her £1. Suppose Igor thinks there's a fifty-fifty chance that he'll succeed. If so, then it's worth it for him to take the bet: the upside is a 50 percent chance of £3, worth £1.50; the downside is a 50 percent chance of losing £1, worth negative £0.50. Igor makes an expected £1 by taking the bet—£1.50 minus £0.50. If his beliefs about his own chances of success are accurate, then if he were to take this bet over and over again, on average he'd make £1 each time.



**Table 2.2. Igor's Decision**

| | Catches the coasters (50% probability) | Fails to catch the coasters (50% probability) | Expected payoff |
|---|---|---|---|
| Take bet | £3 | −£1 | £1 |
| Refuse bet | £0 | £0 | £0 |

Expected value theory is not just useful when gambling. It's crucial whenever we have to take a bet—that is, to make a decision in the face of uncertainty—which is almost all the time. My teenage decisions make this vivid. Before going buildering, I dismissed the possibility of falling and dying as unlikely and therefore not worth worrying about. But that was hugely foolish—not because it was *likely* that I would fall and die, but because it wasn't *sufficiently unlikely*, and dying is so bad that even a small chance is well worth avoiding.

In the face of an uncertain future, humanity often acts like my reckless teenage self. For example, climate change sceptics often point to our uncertainty as a reason for inaction.[26] There's so much we don't know, they claim—we don't know exactly how well climate models predict the amount of warming for a given quantity of emissions, for instance, or just how damaging a certain amount of warming would be for the economy. So we should not waste resources on the problem. But this is a terrible argument. We can grant that there's great uncertainty about what climate change means. But uncertainty cuts both ways. The damage caused by climate change might be less than is typically forecasted, but it might also be considerably *worse*—if, for example, the climate is more sensitive to temperature changes than such forecasts presuppose, or adaptation is harder, or we will emit more carbon dioxide than experts currently predict.

Crucially, the uncertainty around climate change is not symmetric: greater uncertainty should prompt more concern about worst-case outcomes, and this shift is not offset by a higher chance of best-case outcomes, because the worst-case outcomes are worse than the best-case outcomes are good.[27] For example, according to the Intergovernmental Panel on Climate Change, on the medium-low-emissions scenario, the best guess is that we will end

up with around 2.5 degrees Celsius of warming by the end of the century.[28] But this is uncertain. There is a one-in-ten chance that we get 2 degrees or less. But that should not reassure us, because there is also a one-in-ten chance that we get more than 3.5 degrees.[29] Less than 2 degrees would be something of a relief compared to the best-guess estimate, but more than 3.5 degrees would be much worse. The uncertainty gives us *more* reason to worry, not less. It's as if my teenage self, before jumping off a building, had reassured onlookers by saying, "It's OK, I've no idea how far I'll fall!"

Much the same will be true for the issues that I cover in this book. I'm not saying that we should be confident that value lock-in or major catastrophe will occur this century. What I am saying is that their chance of occurring is very real—certainly more than 1 percent, and certainly greater than many everyday risks, like dying in a car crash. When combined with how much is at stake, the expected value of trying to ensure a good future is enormous.

When we're applying the significance, persistence, and contingency framework, we should therefore be thinking about expected significance, expected persistence, and expected contingency.[30] If some change to the world has an 80 percent chance of fizzling out after ten years but a 20 percent chance of lasting for a million years, then its expected persistence is over two hundred thousand years. In general, if some change to the world has at least a reasonable chance of being highly significant, persistent, and contingent, then that can be sufficient for the expected value of that change to be very great indeed.

## Moments of Plasticity

Often, some event can have highly significant, persistent, and contingent effects if there is a period of plasticity, where ideas or events or institutions can take one of many forms, followed by a period of rigidity or ossification. The dynamic is like that of glassblowing: In one period, the glass is still molten and malleable; it can be blown into one of many shapes. After it cools, it becomes rigid, and further change is impossible without remelting.

Plasticity frequently comes after a crisis, like a war. For example, after the end of World War II, Korea was divided along the thirty-eighth parallel. The location of the division was extremely contingent. Colonel Dean Rusk and Charles Bonesteel, two American officers in their midthirties using a



*National Geographic* map, proposed the thirty-eighth parallel because it divided the country roughly in half while keeping Seoul on the American side.[31] They were working on short notice because the United States had to reach an agreement with the Soviet Union before the entire peninsula fell into Soviet hands. No experts on Korea were consulted, and the proposed border cut across several preexisting Korean provinces and geographic features. In fact, the United States was surprised that the Soviets accepted the division; not only did it give Seoul to the United States, but Soviet troops were already in Korea while the closest American forces were still in Okinawa, several hundred miles away.[32] Yet after the division was implemented, it became hard to reverse, and it has since resulted in enormous differences to the fates of those who ended up in each of those two countries. South Koreans live in a strong democracy and are almost thirty times richer on average than they were in 1953. North Koreans live under a totalitarian dictatorship and may be even poorer than they were before the Korean War.[33]

A period of plasticity also commonly occurs when some idea or institution is still new. For example, the US Constitution was written over just four months—a moment of great plasticity—and amended eleven times in its first six years of operation.[34] After that, though, it became more rigid. Between 1804 and 1913, only three amendments were passed, all immediately following the Civil War: they abolished slavery, granted citizenship to African Americans and formerly enslaved people, and prohibited race from influencing the right to vote.[35] Today, the Constitution is again very rigid: it's only been amended once in the last fifty years, and that amendment—to prevent increases in congressional salaries from taking effect until the next term of office—was first proposed in 1789.[36]

This dynamic can hold for the laws and norms relevant to new technologies, too. Following World War II, the international community debated a variety of ways nuclear weapons could be governed.[37] One proposal, put forward by the United States, was the Baruch Plan, according to which the United States would disband its nuclear weapons programme and transfer its bombs to the UN to be destroyed. The UN would then oversee the mining of fissionable materials around the world and inspect other countries to ensure that no one was building nuclear bombs. The USSR countered with the Gromyko Plan, which also proposed universal disarmament. Both of

these plans failed, and it's not clear that either ever had much of a chance. But it was clearly a time of much greater plasticity in nuclear governance than we see now. Today, the idea that the UN could control the mining of uranium seems entirely off the table.

The dynamic of "early plasticity, later rigidity" can hold for new ideas, too. In addition to the books that we now know as the New Testament, a number of other texts were taught by some early Christians.[38] The New Testament books became the core Christian teachings only over the course of the first and second centuries AD and were not cemented until around the end of the fourth century AD.[39]

A final example comes from the history of climate change activism. The effect that carbon dioxide would have on global warming was first quantified in 1896 by Svante Arrhenius; his 1906 estimate of equilibrium climate sensitivity was four degrees, which is only a little higher than modern estimates.[40] And it was knowable, at that time, that we would probably emit dramatically more carbon dioxide in the future: one simply needed to continue extrapolating the trend of exponential economic growth and to recognize the obvious fact that such growth would bring a corresponding increase in energy demand.

In 1958, Frank Capra, director of *It's a Wonderful Life*, made an educational weather documentary, *Unchained Goddess*, which included a warning about climate change: "Even now, man may be unwittingly changing the world's climate through the waste products of his civilisation. Due to our release through factories and automobiles every year of more than six billion tonnes of carbon dioxide, which helps air absorb heat from the sun, our atmosphere seems to be getting warmer. . . . [It's] been calculated that a few degrees rise in the earth's temperature would melt the polar ice caps."[41] Two years earlier, referencing work by Gilbert Plass, the *New York Times* had published an article arguing that carbon dioxide emissions were warming the planet. As with Svante Arrhenius's, Plass's estimate of equilibrium climate sensitivity—3.6 degrees—was strikingly close to the Intergovernmental Panel on Climate Change's current best estimate.[42]

If we had taken action on climate change earlier, we would have been acting on more speculative evidence than we have now. But the issue would also have been much less politically divisive, and change might have been much easier. Bill McKibben, one of the world's leading environmentalists, suggested this, saying in 2019: "Thirty years ago, there were relatively small things we could have done that would have changed the trajectory of this battle—a small price on carbon back then would have yielded a different trajectory, would have put us in a different place. We might not have solved climate change yet because it's a huge problem, but we'd be on the way."[43]



The lesson Bill McKibben takes from the history of climate change activism is that we should pay close attention to new challenges as they arise. He highlights advanced artificial intelligence in particular: "We haven't taken [advanced artificial intelligence] seriously because it doesn't, at the moment, impinge on our day-to-day life. But one of the things that climate change taught me is that things happen fast, like, really fast. And, before you know it, they're out of control. So the time for thinking about them is when there is still some chance of getting a handle on them."[44] He's right. With climate change, we may have missed one moment of plasticity, and we should hope there are more to come. But perhaps we can also learn a more general lesson and respond more rapidly to new challenges—like artificial intelligence, synthetic biology, tensions between the United States and China, the rise of new ideologies, and the potential slowdown in technological progress—as soon as they arise. These are some of the issues I'll cover in the next two parts of this book.

Indeed, over the next two chapters, I'll suggest that the dynamic of "early plasticity, later rigidity" could be true for history as a whole. We are currently in a period where the values that guide civilisation are still malleable, but I'll argue in Chapter 4 that, within the next few centuries, those values could ossify, constraining the course of all future civilisation. If so, then changes we make to today's moral values could have indefinitely long-lasting impacts. Let's turn to this idea, focusing first on the *contingency* of moral change.