

INTRODUCTION

If all goes well, human history is just beginning. Humanity is about two hundred thousand years old. But the Earth will remain habitable for hundreds of millions more—enough time for millions of future generations; enough to end disease, poverty and injustice forever; enough to create heights of flourishing unimaginable today. And if we could learn to reach out further into the cosmos, we could have more time yet: trillions of years, to explore billions of worlds. Such a lifespan places present-day humanity in its earliest infancy. A vast and extraordinary adulthood awaits.

Our view of this potential is easily obscured. The latest scandal draws our outrage; the latest tragedy, our sympathy. Time and space shrink. We forget the scale of the story in which we take part. But there are moments when we remember—when our vision shifts, and our priorities realign. We see a species precariously close to self-destruction, with a future of immense promise hanging in the balance. And which way that balance tips becomes our most urgent public concern.

This book argues that safeguarding humanity's future is the defining challenge of our time. For we stand at a crucial moment in the history of our species. Fueled by technological progress, our power has grown so great that for the first time in humanity's long history, we have the capacity to destroy ourselves—severing our entire future and everything we could become.

Yet humanity's wisdom has grown only falteringly, if at all, and lags dangerously behind. Humanity lacks the maturity, coordination and foresight necessary to avoid making mistakes from which we could never recover. As the gap between our power and our wisdom grows, our future is subject to an ever-increasing level of risk. This situation is unsustainable. So over the next few centuries, humanity will be tested: it will either act

decisively to protect itself and its longterm potential, or, in all likelihood, this will be lost forever.

To survive these challenges and secure our future, we must act now: managing the risks of today, averting those of tomorrow, and becoming the kind of society that will never pose such risks to itself again.

It is only in the last century that humanity's power to threaten its entire future became apparent. One of the most harrowing episodes has just recently come to light. On Saturday, October 27, 1962, a single officer on a Soviet submarine almost started a nuclear war. His name was Valentin Savitsky. He was captain of the submarine B-59—one of four submarines the Soviet Union had sent to support its military operations in Cuba. Each was armed with a secret weapon: a nuclear torpedo with explosive power comparable to the Hiroshima bomb.

It was the height of the Cuban Missile Crisis. Two weeks earlier, US aerial reconnaissance had produced photographic evidence that the Soviet Union was installing nuclear missiles in Cuba, from which they could strike directly at the mainland United States. In response, the US blockaded the seas around Cuba, drew up plans for an invasion and brought its nuclear forces to the unprecedented alert level of DEFCON 2 (“Next step to nuclear war”).

On that Saturday, one of the blockading US warships detected Savitsky's submarine and attempted to force it to the surface by dropping low-explosive depth charges as warning shots. The submarine had been hiding deep underwater for days. It was out of radio contact, so the crew did not know whether war had already broken out. Conditions on board were extremely bad. It was built for the Arctic and its ventilator had broken in the tropical water. The heat inside was unbearable, ranging from 113°F near the torpedo tubes to 140°F in the engine room. Carbon dioxide had built up to dangerous concentrations, and crew members had begun to fall unconscious. Depth charges were exploding right next to the hull. One of the crew later recalled: “It felt like you were sitting in a metal barrel, which somebody is constantly blasting with a sledgehammer.”

Increasingly desperate, Captain Savitsky ordered his crew to prepare their secret weapon:

Maybe the war has already started up there, while we are doing somersaults here. We're going to blast them now! We will die, but we will sink them all—we will not disgrace our Navy!¹

Firing the nuclear weapon required the agreement of the submarine's political officer, who held the other half of the firing key. Despite the lack of authorization by Moscow, the political officer gave his consent.

On any of the other three submarines, this would have sufficed to launch their nuclear weapon. But by the purest luck, submarine B-59 carried the commander of the entire flotilla, Captain Vasili Arkhipov, and so required his additional consent. Arkhipov refused to grant it. Instead, he talked Captain Savitsky down from his rage and convinced him to give up: to surface amidst the US warships and await further orders from Moscow.²

We do not know precisely what would have happened if Arkhipov had granted his consent—or had he simply been stationed on any of the other three submarines. Perhaps Savitsky would not have followed through on his command. What is clear is that we came precariously close to a nuclear strike on the blockading fleet—a strike which would most likely have resulted in nuclear retaliation, then escalation to a full-scale nuclear war (the only kind the US had plans for). Years later, Robert McNamara, Secretary of Defense during the crisis, came to the same conclusion:

No one should believe that had U.S. troops been attacked by nuclear warheads, the U.S. would have refrained from responding with nuclear warheads. Where would it have ended? In utter disaster.³

Ever since the advent of nuclear weapons, humans have been making choices with such stakes. Ours is a world of flawed decision-makers, working with strikingly incomplete information, directing technologies which threaten the entire future of the species. We were lucky, that Saturday in 1962, and have so far avoided catastrophe. But our destructive capabilities continue to grow, and we cannot rely on luck forever.

We need to take decisive steps to end this period of escalating risk and safeguard our future. Fortunately, it is in our power to do so. The greatest risks are caused by human action, and they can be addressed by human

action. Whether humanity survives this era is thus a choice humanity will make. But it is not an easy one. It all depends on how quickly we can come to understand and accept the fresh responsibilities that come with our unprecedented power.

This is a book about *existential risks*—risks that threaten the destruction of humanity’s longterm potential. Extinction is the most obvious way humanity’s entire potential could be destroyed, but there are others. If civilization across the globe were to suffer a truly unrecoverable collapse, that too would destroy our longterm potential. And we shall see that there are dystopian possibilities as well: ways we might get locked into a failed world with no way back.

While this set of risks is diverse, it is also exclusive. So I will have to set aside many important risks that fall short of this bar: our topic is not new dark ages for humanity or the natural world (terrible though they would be), but the permanent destruction of humanity’s potential.

Existential risks present new kinds of challenges. They require us to coordinate globally and intergenerationally, in ways that go beyond what we have achieved so far. And they require foresight rather than trial and error. Since they allow no second chances, we need to build institutions to ensure that across our entire future we never once fall victim to such a catastrophe.

To do justice to this topic, we will have to cover a great deal of ground. Understanding the risks requires delving into physics, biology, earth science and computer science; situating this in the larger story of humanity requires history and anthropology; discerning just how much is at stake requires moral philosophy and economics; and finding solutions requires international relations and political science. Doing this properly requires deep engagement with each of these disciplines, not just cherry-picking expert quotes or studies that support one’s preconceptions. This would be an impossible task for any individual, so I am extremely grateful for the extensive advice and scrutiny of dozens of the world’s leading researchers from across these fields.⁴

This book is ambitious in its aims. Through careful analysis of the potential of humanity and the risks we face, it makes the case that we live

during the most important era of human history. Major risks to our entire future are a new problem, and our thinking has not caught up. So *The Precipice* presents a new ethical perspective: a major reorientation in the way we see the world, and our role in it. In doing so, the book aspires to start closing the gap between our wisdom and power, allowing humanity a clear view of what is at stake, so that we will make the choices necessary to safeguard our future.

I have not always been focused on protecting our longterm future, coming to the topic only reluctantly. I am a philosopher, at Oxford University, specializing in ethics. My earlier work was rooted in the more tangible concerns of global health and global poverty—in how we could best help the worst off. When coming to grips with these issues I felt the need to take my work in ethics beyond the ivory tower. I began advising the World Health Organization, World Bank and UK government on the ethics of global health. And finding that my own money could do hundreds of times as much good for those in poverty as it could do for me, I made a lifelong pledge to donate at least a tenth of all I earn to help them.⁵ I founded a society, *Giving What We Can*, for those who wanted to join me, and was heartened to see thousands of people come together to pledge more than £1 billion over our lifetimes to the most effective charities we know of, working on the most important causes. Together, we've already been able to transform the lives of tens of thousands of people.⁶ And because there are many other ways beyond our donations in which we can help fashion a better world, I helped start a wider movement, known as *effective altruism*, in which people aspire to use evidence and reason to do as much good as possible.

Since there is so much work to be done to fix the needless suffering in our present, I was slow to turn to the future. It was so much less visceral; so much more abstract. Could it really be as urgent a problem as suffering now? As I reflected on the evidence and ideas that would culminate in this book, I came to realize that the risks to humanity's future are just as real and just as urgent—yet even more neglected. And that the people of the future may be even more powerless to protect themselves from the risks we impose than the dispossessed of our own time.

Addressing these risks has now become the central focus of my work: both researching the challenges we face, and advising groups such as the

UK Prime Minister's Office, the World Economic Forum and DeepMind on how they can best address these challenges. Over time, I've seen a growing recognition of these risks, and of the need for concerted action.

To allow this book to reach a diverse readership, I've been ruthless in stripping out the jargon, needless technical detail and defensive qualifications typical of academic writing (my own included). Readers hungry for further technical detail or qualifications can delve into the many endnotes and appendices, written with them in mind.⁷

I have tried especially hard to examine the evidence and arguments carefully and even-handedly, making sure to present the key points even if they cut against my narrative. For it is of the utmost importance to get to the truth of these matters—humanity's attention is scarce and precious, and must not be wasted on flawed narratives or ideas⁸.

Each chapter of *The Precipice* illuminates the central questions from a different angle. Part One (The Stakes) starts with a bird's-eye view of our unique moment in history, then examines why it warrants such urgent moral concern. Part Two (The Risks) delves into the science of the risks facing humanity, both from nature and from ourselves, showing that while some have been overstated, there is real risk and it is growing. So Part Three (The Path Forward) develops tools for understanding how these risks compare and combine, and new strategies for addressing them. I close with a vision of our future: of what we could achieve were we to succeed.

This book is not just a familiar story of the perils of climate change or nuclear war. These risks that first awoke us to the possibilities of destroying ourselves are just the beginning. There are emerging risks, such as those arising from biotechnology and advanced artificial intelligence, that may pose much greater risk to humanity in the coming century.

Finally, this is not a pessimistic book. It does not present an inevitable arc of history culminating in our destruction. It is not a morality tale about our technological hubris and resulting fall. Far from it. The central claim is that there are real risks to our future, but that our choices can still make all the difference. I believe we are up to the task: that through our choices we can pull back from the precipice and, in time, create a future of astonishing value—with a richness of which we can barely dream, made possible by innovations we are yet to conceive. Indeed, my deep optimism about humanity's future is core to my motivation in writing this book. Our

potential is vast. We have so much to protect.

2

EXISTENTIAL RISK

The crucial role we fill, as moral beings, is as members of a cross-generational community, a community of beings who look before and after, who interpret the past in light of the present, who see the future as growing out of the past, who see themselves as members of enduring families, nations, cultures, traditions.

—Annette Baier¹

We have seen how the long arc of human history has brought us to a uniquely important time in our story—a period where our entire future hangs in the balance. And we have seen a little of what might lie beyond, if only we can overcome the risks.

It is now time to think more deeply about what is at stake; why safeguarding humanity through this time is so important. To do so we first need to clarify the idea of existential risk. What exactly *is* existential risk? How does it relate to more familiar ideas of extinction or the collapse of civilization?

We can then ask what it is about these risks that compels our urgent concern. The chief reason, in my view, is that we would lose our entire future: everything humanity could be and everything we could achieve. But that is not all. The case that it is crucial to safeguard our future draws support from a wide range of moral traditions and foundations. Existential risks also threaten to destroy our present, and to betray our past. They test civilization's virtues, and threaten to remove what may be the most complex and significant part of the universe.

If we take any of these reasons seriously, we have a lot of work to do to

protect our future. For existential risk is greatly neglected: by government, by academia, by civil society. We will see why this has been the case, and why there is good reason to suspect this will change.

UNDERSTANDING EXISTENTIAL RISK

Humanity's future is ripe with possibility. We have achieved a rich understanding of the world we inhabit and a level of health and prosperity of which our ancestors could only dream. We have begun to explore the other worlds in the heavens above us, and to create virtual worlds completely beyond our ancestors' comprehension. We know of almost no limits to what we might ultimately achieve.

Human extinction would foreclose our future. It would destroy our potential. It would eliminate all possibilities but one: a world bereft of human flourishing. Extinction would bring about this failed world and lock it in forever—there would be no coming back.

The philosopher Nick Bostrom showed that extinction is not the only way this could happen: there are other catastrophic outcomes in which we lose not just the present, but all our potential for the future.²

Consider a world in ruins: an immense catastrophe has triggered a global collapse of civilization, reducing humanity to a pre-agricultural state. During this catastrophe, the Earth's environment was damaged so severely that it has become impossible for the survivors to ever re-establish civilization. Even if such a catastrophe did not cause our extinction, it would have a similar effect on our future. The vast realm of futures currently open to us would have collapsed to a narrow range of meager options. We would have a failed world with no way back.

Or consider a world in chains: in a future reminiscent of George Orwell's *Nineteen Eighty-Four*, the entire world has become locked under the rule of an oppressive totalitarian regime, determined to perpetuate itself. Through powerful, technologically enabled indoctrination, surveillance and enforcement, it has become impossible for even a handful of dissidents to find each other, let alone stage an uprising. With everyone on Earth living under such rule, the regime is stable from threats, internal and external. If such a regime could be maintained indefinitely, then descent into this

totalitarian future would also have much in common with extinction: just a narrow range of terrible futures remaining, and no way out.

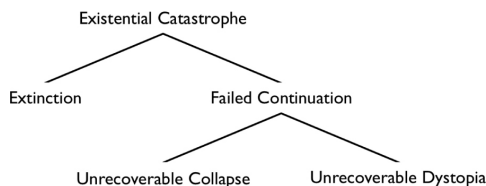


FIGURE 2.1 A classification of existential catastrophes by the kind of outcome that gets locked in.

Following Bostrom, I shall call these “existential catastrophes,” defining them as follows:³

An *existential catastrophe* is the destruction of humanity’s longterm potential.

An *existential risk* is a risk that threatens the destruction of humanity’s longterm potential.

These definitions capture the idea that the outcome of an existential catastrophe is both dismal and irrevocable. We will not just fail to fulfill our potential, but this very potential itself will be permanently lost. While I want to keep the official definitions succinct, there are several areas that warrant clarification.

First, I am understanding *humanity’s longterm potential* in terms of the set of all possible futures that remain open to us.⁴ This is an expansive idea of possibility, including everything that humanity could eventually achieve, even if we have yet to invent the means of achieving it.⁵ But it follows that while our choices can lock things in, closing off possibilities, they can’t open up new ones. So any reduction in humanity’s potential should be understood as permanent. The challenge of our time is to *preserve* our vast potential, and to *protect* it against the risk of future destruction. The ultimate purpose is to allow our descendants to *fulfill* our potential, realizing one of the best possible futures open to us.

While it may seem abstract at this scale, this is really a familiar idea that

we encounter every day. Consider a child with high longterm potential: with futures open to her in which she leads a great life. It is important that her potential is preserved: that her best futures aren't cut off due to accident, trauma or lack of education. It is important that her potential is protected: that we build in safeguards to make such a loss of potential extremely unlikely. And it is important that she ultimately fulfills her potential: that she ends up taking one of the best paths open to her. So too for humanity.

Existential risks threaten the destruction of humanity's potential. This includes cases where this destruction is complete (such as extinction) and where it is nearly complete, such as a permanent collapse of civilization in which the possibility for some very minor types of flourishing remain, or where there remains some remote chance of recovery.⁶ I leave the thresholds vague, but it should be understood that in any existential catastrophe the greater part of our potential is gone and very little remains.⁷

Second, my focus on humanity in the definitions is not supposed to exclude considerations of the value of the environment, other animals, successors to *Homo sapiens*, or creatures elsewhere in the cosmos. It is not that I think only humans count. Instead, it is that humans are the only beings we know of that are responsive to moral reasons and moral argument—the beings who can examine the world and decide to do what is best. If we fail, that upward force, that capacity to push toward what is best or what is just, will vanish from the world.

Our potential is a matter of what humanity can achieve through the combined actions of each and every human. The value of our actions will stem in part from what we do to and for humans, but it will depend on the effects of our actions on non-humans too. If we somehow give rise to new kinds of moral agents in the future, the term “humanity” in my definition should be taken to include them.

My focus on humanity prevents threats to a single country or culture from counting as existential risks. There is a similar term that gets used this way—when people say that something is “an existential threat to this country.” Setting aside the fact that these claims are usually hyperbole, they are expressing a similar idea: that something threatens to permanently destroy the longterm potential of a country or culture.⁸ However, it is very important to keep talk of an “existential risk” (without any explicit restriction to a group) to apply only to threats against the whole of

humanity.

Third, any notion of risk must involve some kind of probability. What kind is involved in existential risk? Understanding the probability in terms of objective long-run frequencies won't work, as the existential catastrophes we are concerned with can only ever happen once, and will always be unprecedented until the moment it is too late. We can't say the probability of an existential catastrophe is precisely zero just because it hasn't happened *yet*.

Situations like these require an evidential sense of probability, which describes the appropriate degree of belief we should have on the basis of the available information. This is the familiar type of probability used in courtrooms, banks and betting shops. When I speak of the probability of an existential catastrophe, I will mean the credence humanity should have that it will occur, in light of our best evidence.⁹

There are many utterly terrible outcomes that do not count as existential catastrophes.

One way this could happen is if there were no single precipitous event, but a multitude of smaller failures. This is because I take on the usual sense of catastrophe as a single, decisive event, rather than any combination of events that is bad in sum. If we were to squander our future simply by continually treating each other badly, or by never getting around to doing anything great, this could be just as bad an outcome but wouldn't have come about via a catastrophe.

Alternatively, there might be a single catastrophe, but one that leaves open some way for humanity to eventually recover. From our own vantage, looking out to the next few generations, this may appear equally bleak. But a thousand years hence it may be considered just one of several dark episodes in the human story. A true existential catastrophe must by its very nature be the decisive moment of human history—the point where we failed.

Even catastrophes large enough to bring about the global collapse of civilization may fall short of being existential catastrophes. While colloquially referred to as “the end of the world,” a global collapse of civilization need not be the end of the human story. It has the required

severity, but may not be permanent or irrevocable.

In this book, I shall use the term *civilization collapse* quite literally, to refer to an outcome where humanity across the globe loses civilization (at least temporarily), being reduced to a pre-agricultural way of life. The term is often used loosely to refer merely to a massive breakdown of order, the loss of modern technology, or an end to our culture. But I am talking about a world without writing, cities, law, or any of the other trappings of civilization.

This would be a very severe disaster and extremely hard to trigger. For all the historical pressures on civilizations, never once has this happened—not even on the scale of a continent.¹⁰ The fact that Europe survived losing 25 to 50 percent of its population in the Black Death, while keeping civilization firmly intact, suggests that triggering the collapse of civilization would require more than 50 percent fatality in every region of the world.¹¹

Even if civilization did collapse, it is likely that it could be re-established. As we have seen, civilization has already been independently established at least seven times by isolated peoples.¹² While one might think resource depletion could make this harder, it is more likely that it has become substantially easier. Most disasters short of human extinction would leave our domesticated animals and plants, as well as copious material resources in the ruins of our cities—it is much easier to re-forged iron from old railings than to smelt it from ore. Even expendable resources such as coal would be much easier to access, via abandoned reserves and mines, than they ever were in the eighteenth century.¹³ Moreover, evidence that civilization is possible, and the tools and knowledge to help rebuild, would be scattered across the world.

There are, however, two close connections between the collapse of civilization and existential risk. First, a collapse would count as an existential catastrophe if it were unrecoverable. For example, it is conceivable that some form of extreme climate change or engineered plague might make the planet so inhospitable that humanity would be irrevocably reduced to scattered foragers.¹⁴ And second, a global collapse of civilization could increase the chance of extinction, by leaving us more vulnerable to subsequent catastrophe.

One way a collapse could lead to extinction is if the population of the largest remaining group fell below the *minimum viable population*—the

level needed for a population to survive. There is no precise figure for this, as it is usually defined probabilistically and depends on many details of the situation: where the population is, what technology they have access to, the sort of catastrophe they have suffered. Estimates range from hundreds of people up to tens of thousands.¹⁵ If a catastrophe directly reduces human population to below these levels, it will be more useful to classify it as a direct extinction event, rather than an unrecoverable collapse. And I expect that this will be one of the more common pathways to extinction.

We rarely think seriously about risks to humanity's entire potential. We encounter them mostly in action films, where our emotional reactions are dulled by their overuse as an easy way to heighten the drama.¹⁶ Or we see them in online lists of "ten ways the world could end," aimed primarily to thrill and entertain. Since the end of the Cold War, we rarely encounter sober discussions by our leading thinkers on what extinction would mean for us, our cultures or humanity.¹⁷ And so in casual contexts people are sometimes flippant about the prospect of human extinction.

But when a risk is made vivid and credible—when it is clear that billions of lives and all future generations are actually on the line—the importance of protecting humanity's longterm potential is not, for most people, controversial. If we learned that a large asteroid was heading toward Earth, posing a greater than 10 percent chance of human extinction later this century, there would be little debate about whether to make serious efforts to build a deflection system, or to ignore the issue and run the risk. To the contrary, responding to the threat would immediately become one of the world's top priorities. Thus our lack of concern about these threats is much more to do with not yet believing that there are such threats, than it is about seriously doubting the immensity of the stakes.

Yet it is important to spend a little while trying to understand more clearly the different sources of this importance. Such an understanding can buttress feeling and inspire action; it can bring to light new considerations; and it can aid in decisions about how to set our priorities.

[LOOKING TO THE PRESENT](#)

Not all existential catastrophes involve human extinction, and not all methods of extinction involve pain or untimely death. For example, it is theoretically possible that we could all simply decide not to reproduce. This could destroy our potential without, let us suppose, causing any suffering. But the existential risks we actually face are not so peaceful. Rather, they are obviously horrific by the most familiar moral standards.

If, over the coming century, humanity is destroyed in a nuclear winter, or an engineered pandemic, or a catastrophic war involving some new technology, then seven billion lives would be cut short—including, perhaps, your own life, or the lives of those you love. Many would likely die in agony—starving, or burning, or ravaged by disease.

The moral case for preventing such horror needs little elaboration. Humanity has seen catastrophes before, on smaller scales: thousands, or millions, of human lives destroyed. We know how tremendously important it is to prevent such disasters. At such a scale, we lose our ability to fully comprehend the magnitude of what is lost, but even then the numbers provide a guide to the moral stakes.¹⁸ Other things being equal, millions of deaths must be much worse than thousands of deaths; and billions, much worse than millions. Even measured just in terms of lives cut short, human extinction would easily be the worst event in our long history.

LOOKING TO OUR FUTURE

But an existential catastrophe is not just a catastrophe that destroys a particularly large number of lives. It destroys our potential.

My mentor, Derek Parfit, asked us to imagine a devastating nuclear war killing 99 percent of the world's people.¹⁹ A war that would leave behind a dark age lasting centuries, before the survivors could eventually rebuild civilization to its former heights; humbled, scarred—but undefeated.

Now compare this with a war killing a full 100 percent of the world's people. This second war would be worse, of course, but how much worse? Either war would be the worst catastrophe in history. Either would kill billions. The second war would involve tens of millions of additional deaths, and so would be worse for this reason. But there is another, far more significant difference between the two wars. Both wars kill billions of

humans; but the second war kills humanity. Both wars destroy our present; but the second war destroys our future.

It is this qualitative difference in what is lost with that last percent that makes existential catastrophes unique, and that makes reducing the risk of existential catastrophe uniquely important.^{[20](#)}

In expectation, almost all humans who will ever live have yet to be born. Absent catastrophe, most generations are future generations. As the writer Jonathan Schell put it:

The procession of generations that extends onwards from our present leads far, far beyond the line of our sight, and, compared with these stretches of human time, which exceed the whole history of the earth up to now, our brief civilized moment is almost infinitesimal. Yet we threaten, in the name of our transient aims and fallible convictions, to foreclose it all. If our species does destroy itself, it will be a death in the cradle—a case of infant mortality.^{[21](#)}

And because, in expectation, almost all of humanity's life lies in the future, almost everything of value lies in the future as well: almost all the flourishing; almost all the beauty; our greatest achievements; our most just societies; our most profound discoveries.^{[22](#)} We can continue our progress on prosperity, health, justice, freedom and moral thought. We can create a world of wellbeing and flourishing that challenges our capacity to imagine. And if we protect that world from catastrophe, it could last millions of centuries. This is our potential—what we could achieve if we pass the Precipice and continue striving for a better world.

It is this view of the future—the immense value of humanity's potential—that most persuades me to focus my energies on reducing existential risk. When I think of the millions of future generations yet to come, the importance of protecting humanity's future is clear to me. To risk destroying this future, for the sake of some advantage limited only to the present, seems to me profoundly parochial and dangerously short-sighted. Such neglect privileges a tiny sliver of our story over the grand sweep of the whole; it privileges a tiny minority of humans over the overwhelming majority yet to be born; it privileges this particular century over the

millions, or maybe billions, yet to come.²³

To see why this would be wrong, consider an analogy with distance. A person does not matter less, the further away from you they are in space. It matters just as much if my wife gets sick while she is away at a conference in Kenya as if she gets sick while home with me in Oxford. And the welfare of strangers in Kenya matters just as much as the welfare of strangers in Oxford. Of course, we may have special duties to some individuals—to family; to members of the same community—but it is never spatial distance, in itself, that determines these differences in our obligations. Recognizing that people matter equally, regardless of their geographic location, is a crucial form of moral progress, and one that we could do much more to integrate into our policies and our philanthropy.

People matter equally regardless of their temporal location too. Our lives matter just as much as those lived thousands of years ago, or those a thousand years hence.²⁴ Just as it would be wrong to think that other people matter less the further they are from you in space, so it is to think they matter less the further away from you they are in time. The value of their happiness, and the horror of their suffering, is undiminished.

Recognizing that people matter equally, wherever they are in time, is a crucial next step in the ongoing story of humanity's moral progress. Many of us recognize this equality to some extent already. We know it is wrong to make future generations worse off in order to secure lesser benefits for ourselves. And if asked, we would agree that people now don't objectively matter more than people in the future. But we assume that this leaves most of our priorities unaltered. For example, thinking that long-run effects of our choices quickly disappear; that they are so uncertain that the good cancels the bad; or that people in the future will be much better situated to help themselves.²⁵

But the possibility of preventable existential risks in our lifetimes shows that there are issues where our actions can have sustained positive effects over the whole longterm future, and where we are the only generation in a position to produce those effects.²⁶ So the view that people in the future matter just as much as us has deep practical implications. We have a long way to go if we are to understand these and integrate them fully into our moral thinking.

Considerations like these suggest an ethic we might call *longtermism*,

which is especially concerned with the impacts of our actions upon the longterm future.²⁷ It takes seriously the fact that our own generation is but one page in a much longer story, and that our most important role may be how we shape—or fail to shape—that story. Working to safeguard humanity’s potential is one avenue for such a lasting impact and there may be others too.²⁸

One doesn’t have to approach existential risk from this direction—there is already a strong moral case just from the immediate effects—but a longtermist ethic is nevertheless especially well suited to grappling with existential risk. For longtermism is animated by a moral re-orientation toward the vast future that existential risks threaten to foreclose.

Of course, there are complexities.

When economists evaluate future benefits, they use a method called discounting, which dampens (“discounts”) benefits based on how far away they are in time. If one took a commonly used discount rate of 5 percent per year and applied it to our future, there would be strikingly little value left. Applied naïvely, this discount rate would suggest our entire future is worth only about twenty times as much as the coming year, and that the period from 2100 to eternity is worth less than the coming year. Does this call into question the idea that our future is extremely valuable?

No. Results like this arise only from an incorrect application of the economic methods. When the subtleties of the problem are taken into account and discounting is correctly applied, the future is accorded an extremely high value. The mathematical details would take us too far afield, but for now it suffices to note that discounting human wellbeing (as opposed to instrumental goods such as money), purely on the basis of distance away from us in time, is deeply implausible—especially over the long time periods we are discussing. It implies, for example, that if you can save one person from a headache in a million years’ time, or a billion people from torture in two million years, you should save the one from a headache.²⁹ A full explanation of why economic discounting does not trivialize the value of the longterm future can be found in Appendix A.

Some philosophers question the value of protecting our longterm future for quite a different reason. They note that the timing of the benefits is not

the only unusual feature of this case. If we save humanity from extinction, that will change the number of people who will ever live. This brings up ethical issues that don't arise when simply saving the lives of existing people. Some of the more extreme approaches to this relatively new field of "population ethics" imply that there is no reason to avoid extinction stemming from considerations of future generations—it just doesn't matter whether these future people come into being or not.

A full treatment of these matters would take too long and be of interest only to a few, so I reserve the detailed discussion for Appendix B. To briefly summarize: I do not find these views very plausible, either. They struggle to capture our reasons to care about whether we make future lives worse by polluting the planet, or changing the climate, and to explain why we have strong reasons to prevent terrible lives from existing in the future. And all but the most implausible of these views agree with the immense importance of saving future generations from other kinds of existential catastrophe, such as the irrevocable collapse of civilization. Since most things that threaten extinction threaten such a collapse too, there is not much practical difference. That said, the issues are complex, and I encourage interested readers to consult the appendix for details.

There is one other objection I want to touch on. When I was younger, I sometimes took comfort in the idea that perhaps the outright destruction of humanity would not be bad at all. There would be no people to suffer or grieve. There would be no *badness* at those future times, so how could the destruction be bad? And if the existence of humanity was somehow essential to judgments of right and wrong, good and bad, then perhaps such concepts would fail to apply at all in the stillness that followed.

But I now see that this is no better than the old argument by the philosopher Epicurus that your death cannot be bad for you, since you are not there to experience it. What this neglects is that if I step out into the traffic and die, my life as a whole will be shorter and thereby worse: not by having more that is bad, but by containing less of everything that makes life good. That is why I shouldn't do it. While Epicurus's argument may provide consolation in times of grief or fear, it is not fit to be a guide for action, and no one treats it so. Imagine a government using it as the basis for our policies on safety or healthcare—or for our laws on murder.

If a catastrophe this century were to cause our extinction, then

humanity's life would be shorter and thereby worse.³⁰ Given that we may just be in our infancy, it would be much shorter; much worse. Even if there were no one remaining to judge this as a tragedy, we can rightly judge it so from here. Just as we can judge events in other places, so we can judge events in other times.³¹ And if these judgments are correct now, they shall remain correct when we are no more. I wouldn't blame people who, in humanity's final hours, found consolation in such Epicurean arguments. But the length and quality of humanity's life is still ours to decide, and we must own this responsibility.³²

These are not the only possible objections. Yet we need not resolve every philosophical issue about the value of the future in order to decide whether humanity's potential is worth protecting. For the idea that it would be a matter of relative indifference whether humanity goes extinct, or whether we flourish for billions of years, is, on its face, profoundly implausible. In this sense, any theory that denies it should be subject to significant skepticism.³³

What's more, the future is not the only moral lens through which to view existential catastrophe. It is the one that grips me most, and that most persuades me to devote my time and energy to this issue, but there are other lenses, drawing on other moral traditions. So let us briefly explore how concern about existential risk could also spring from considerations of our past, our character and our cosmic significance. And thus how people with many different understandings of morality could all end up at this common conclusion.

LOOKING TO OUR PAST

We are not the first generation. Our cultures, institutions and norms; our knowledge, technology and prosperity; these were gradually built up by our ancestors, over the course of ten thousand generations. In the last chapter we saw how humanity's remarkable success has relied on our capacity for intergenerational cooperation: inheriting from our parents, making some small improvements of our own, and passing it all down to our children. Without this cooperation we would have no houses or farms, we would have no traditions of dance or song, no writing, no nations.³⁴

This idea was beautifully expressed by the conservative political theorist Edmund Burke. In 1790 he wrote of society:

It is a partnership in all science; a partnership in all art; a partnership in every virtue, and in all perfection. As the ends of such a partnership cannot be obtained except in many generations, it becomes a partnership not only between those who are living, but between those who are living, those who are dead, and those who are to be born.^{[35](#)}

This might give us reasons to safeguard humanity that are grounded in our past—obligations to our grandparents, as well as our grandchildren.

Our ancestors set in motion great projects for humanity that are too big for any single generation to achieve. Projects such as bringing an end to war, forging a just world and understanding our universe. In the year 65 CE, Seneca the Younger explicitly set out such a vast intergenerational project:

The time will come when diligent research over long periods will bring to light things which now lie hidden. A single lifetime, even though entirely devoted to the sky, would not be enough for the investigation of so vast a subject... And so this knowledge will be unfolded only through long successive ages. There will come a time when our descendants will be amazed that we did not know things that are so plain to them... Let us be satisfied with what we have found out, and let our descendants also contribute something to the truth... Many discoveries are reserved for ages still to come, when memory of us will have been effaced.^{[36](#)}

It is astounding to be spoken to so directly across such a gulf of time, and to see this 2,000-year plan continue to unfold.^{[37](#)}

A human, or an entire generation, cannot complete such grand projects. But humanity can. We work together, each generation making a little progress while building up the capacities, resources and institutions to empower future generations to take the next step.

Indeed, when I think of the unbroken chain of generations leading to our

time and of everything they have built for us, I am humbled. I am overwhelmed with gratitude; shocked by the enormity of the inheritance and at the impossibility of returning even the smallest fraction of the favor. Because a hundred billion of the people to whom I owe everything are gone forever, and because what they created is so much larger than my life, than my entire generation.

The same is true at the personal level. In the months after my daughter was born, the magnitude of everything my parents did for me was fully revealed. I was shocked. I told them; thanked them; apologized for the impossibility of ever repaying them. And they smiled, telling me that this wasn't how it worked—that one doesn't repay one's parents. One passes it on.

My parents aren't philosophers. But their remarks suggest another way in which the past could ground our duties to the future. Because the arrow of time makes it so much easier to help people who come after you than people who come before, the best way of understanding the partnership of the generations may be asymmetrical, with duties all flowing forwards in time—paying it forwards. On this view, our duties to future generations may thus be grounded in the work our ancestors did for us when we were future generations.³⁸

So if we drop the baton, succumbing to an existential catastrophe, we would fail our ancestors in a multitude of ways. We would fail to achieve the dreams they hoped for; we would betray the trust they placed in us, their heirs; and we would fail in any duty we had to pay forward the work they did for us. To neglect existential risk might thus be to wrong not only the people of the future, but the people of the past.

It would also be to risk the destruction of everything of value from the past we might have reason to preserve.³⁹ Some philosophers have suggested that the right way to respond to some valuable things is not to promote them, but to protect or preserve them; to cherish or revere them.⁴⁰ We often treat the value of cultural traditions in this way. We see indigenous languages and ways of life under threat—perhaps to be lost forever to this world—and we are filled with a desire to preserve them, and protect them from future threats.

Someone who saw the value of humanity in this light may not be so moved by the loss of what could have been. But they would still be

horrified by extinction: the ruin of every cathedral and temple, the erasure of every poem in every tongue, the final and permanent destruction of every cultural tradition the Earth has known. In the face of serious threats of extinction, or of a permanent collapse of civilization, a tradition rooted in preserving or cherishing the richness of humanity would also cry out for action.⁴¹

Finally, we might have duties to the future arising from the flaws of the past. For we might be able to make up for some of our past wrongs. If we failed now, we could never fulfill any duties we might have to repair the damage we have done to the Earth's environment—cleaning up our pollution and waste; restoring the climate to its pre-industrial state; returning ecosystems to their vanished glory. Or consider that some of the greatest injustices have been inflicted not by individuals upon individuals, but by groups upon groups: systematic persecution, stolen lands, genocides. We may have duties to properly acknowledge and memorialize these wrongs; to confront the acts of our past. And there may yet be ways for the beneficiaries of these acts to partly remedy them or atone for them. Suffering an existential catastrophe would remove any last chance to do so.

CIVILIZATIONAL VIRTUES

If we play our cards right, humanity is at an early stage of life: still in our adolescence; looking forward to a remarkable adulthood. Like an adolescent, we are rapidly coming into our full power and are impatient to flex our muscles, to try out every new capability the moment we acquire it. We show little regard for our future. Sure, we sometimes talk about the “long term,” but by this we usually mean the next decade or two. A long time for a human; a moment for humanity.

Like the adolescent, humanity has no need to plan out the details of the rest of its life. But it does need to make plans that bear in mind the duration and broad shape of that future. Otherwise we cannot hope to know which risks are worth taking, and which skills we need to develop to help us fulfill our potential.

Like many adolescents, humanity is impatient and imprudent; sometimes shockingly so. At times this stems from an inability to

appropriately weigh our short-term gains against our longterm interests. More commonly, it is because we completely neglect our longterm future, not even considering it in our decision-making. And like the adolescent, we often stumble straight into risks without making any kind of conscious decision at all.

This analogy provides us with another lens through which to assess our behavior. Rather than looking at the morality of an individual human's actions as they bear on others, we can address the dispositions and character of humanity as a whole and how these help or undercut its own chances of flourishing. When we look at humanity itself as a group agent, comprising all of us over all time, we can gain insight into the systematic strengths or weaknesses in humanity's ability to achieve flourishing. These are virtues and vices at the largest scale—what we could call *civilizational virtues* and vices. One could treat these as having a fundamental moral significance, or simply as a useful way of diagnosing important weakness in our character and suggesting remedies.

Not all virtues need make sense on this level, but many do. Our lack of regard for risks to our entire future is a deficiency of prudence. When we put the interests of our current generation far above those of the generations to follow, we display our lack of patience.⁴² When we recognize the importance of our future yet still fail to prioritize it, it is a failure of self-discipline. When a backward step makes us give up on our future—or assume it to be worthless—we show a lack of hope and perseverance, as well as a lack of responsibility for our own actions.⁴³

In his celebrated account of virtue, Aristotle suggested that our virtues are governed and guided by a form of practical wisdom. This fits well with the idea of civilizational virtues too. For as our power continues to grow, our practical wisdom needs to grow with it.

COSMIC SIGNIFICANCE

Whether we are alone in the universe is one of the greatest remaining mysteries of science.⁴⁴ Eminent astronomers such as Martin Rees, Max Tegmark and Carl Sagan have reasoned that if we are alone, our survival and our actions might take on a cosmic significance.⁴⁵ While we are

certainly smaller than the galaxies and stars, less spectacular than supernovae or black holes, we may yet be one of the most rare and precious parts of the cosmos.⁴⁶ The nature of such significance would depend on the ways in which we are unique.

If we are the only moral agents that will ever arise in our universe—the only beings capable of making choices on the grounds of what is right and wrong—then responsibility for the history of the universe is entirely *on us*. This is the only chance ever to shape the universe toward what is right, what is just, what is best for all. If we fail, then the potential not just of humanity, but of all moral action, will have been irrevocably squandered.

THE PERSPECTIVE OF HUMANITY

Seeing our predicament from the perspective of humanity is a major theme of this book. Ethics is most commonly addressed from the individual perspective: what should *I* do? Occasionally, it is considered from the perspective of a group or nation, or even (more recently) from the global perspective of everyone alive today. Understanding what the group should do can help its members see the parts they need to play.

We shall sometimes take this a step further, exploring ethics from the perspective of humanity.⁴⁷ Not just our present generation, but humanity over deep time: reflecting on what we achieved in the last 10,000 generations and what we may be able to achieve in the eons to come.

This perspective allows us to see how our own time fits into the greater story, and how much is at stake. It changes the way we see the world and our role in it, shifting our attention from things that affect the fleeting present, to those that could make fundamental alterations to the shape of the longterm future. What matters most for humanity? And what part in this plan should our generation play? What part should I play?⁴⁸

Of course, humanity is not an individual. But it is often useful for us to think about groups as agents, gaining insights by talking about the beliefs, desires and intentions of teams, companies or nations.

Consider how often we speak of a company's strategy, a nation's interests, or even what a country is hoping to achieve with its latest gambit. Such mental states are usually less coherent than those of individuals, as there can be internal tensions between the individuals that comprise the group. But individuals too have their own ambivalence or inner inconsistency, and the idea of "group agents" has proved essential to anyone trying to understand the business world or international landscape.

Applying this perspective to humanity as a whole is increasingly useful and important. Humanity was splintered into isolated peoples for nearly the entire time since civilization began. Only recently have we found each other across the seas and started forming a single global civilization. Only recently have we discovered the length and shape of our long history, or the true potential of our future. And only recently have we faced significant threats that require global coordination.

We shouldn't always take this perspective. Many moral challenges operate at the personal level, or the level of smaller groups. And even when it comes to the big-picture questions, it is sometimes more important to focus on the ways in which humanity is divided: on our differing power or responsibility. But just as we've seen the value of occasionally adopting a global perspective, so too is it important to sometimes step further back and take the perspective of humanity.

The idea of civilizational virtues is just one example of explicitly adopting this perspective. In [Chapter 7](#), we shall do so again, considering grand strategy for humanity. And even when we are looking at our own generation's responsibilities or what we each should do, this will be illuminated by the big picture of humanity across the eons.

Alternatively, if we are the only beings capable of wondering about the universe, then we might have additional reason to seek such understanding. For it would only be through us that a part of the universe could come to fully understand the laws that govern the whole.

And if Earth is the only place in the universe that will give rise to life,

then all life on Earth would have a key significance. Earth would be the only place where there was so much complexity in each drop of water, the only place where anything lived and died, the only place where anything felt, or thought, or loved. And humanity would be the only form of life capable of stewarding life itself, protecting it from natural catastrophes and, eventually, taking it to flourish throughout the cosmos.

UNCERTAINTY

So we could understand the importance of existential risk in terms of our present, our future, our past, our character or our cosmic significance. I am most confident in the considerations grounded in the value of our present and our future, but the availability of other lenses shows the robustness of the case for concern: it doesn't rely on any single school of moral thought, but springs naturally from a great many. While each avenue may suggest a different strength and nature of concern, together they provide a wide base of support for the idea that avoiding existential catastrophe is of grave moral importance.

I'm sure many readers are convinced by this point, but a few will still harbor doubts. I have sympathy, because I too am not *completely* certain. This uncertainty comes in two parts. The first is the everyday kind of uncertainty: uncertainty about what will happen in the future. Might the evidence for humanity's vast potential be misleading? The second is moral uncertainty: uncertainty about the nature of our ethical commitments.⁴⁹ Might I be mistaken about the strength of our obligations to future generations?

However, the case for making existential risk a global priority does not require certainty, for the stakes aren't balanced. If we make serious investments to protect humanity when we had no real duty to do so, we would err, wasting resources we could have spent on other noble causes. But if we neglect our future when we had a real duty to protect it, we would do something far worse—failing forever in what could well be our most important duty. So long as we find the case for safeguarding our future quite plausible, it would be extremely reckless to neglect it.⁵⁰

Even if someone were so pessimistic about the future as to think it

negative in expectation—that the heights we might reach are more than matched by the depths to which we might sink—there is *still* good reason to protect our potential.⁵¹ For one thing, some existential catastrophes (such as permanent global totalitarianism) would remain uncontroversially terrible and thus worthy of our attention. But there is a deeper reason too. In this case there would be immense value of information in finding out more about whether our future will be positive or negative. By far the best strategy would be to protect humanity until we have a much more informed position on this crucial question.⁵²

And it is not just regarding the value of the future that our descendants will be better informed. At present we are still more generally inexperienced. We have little practice at the complexities of managing a global civilization, or a planet. Our view of the future is still clouded by ignorance and distorted by bias. But our descendants, if all goes well, will be far wiser than we are. They will have had time to understand much more deeply the nature of our condition; they will draw strength and insight from a more just, skillful and mature civilization; and their choices, in general, will reflect a fuller understanding of what is at stake when they choose. We in the present day, at what may be the very start of history, would therefore do well to be humble, to leave our options open, and to ensure our descendants have a chance to see more clearly, and choose more wisely, than we can today.⁵³

OUR NEGLECT OF EXISTENTIAL RISKS

The world is just waking up to the importance of existential risk. We have begun work on evaluating and evading the most significant threats, but have yet to scale this up in proportion to the significance of the problems. Seen in the context of the overall distribution of global resources, existential risk is sorely neglected.

Consider the possibility of engineered pandemics, which we shall soon see to be one of the largest risks facing humanity. The international body responsible for the continued prohibition of bioweapons (the Biological Weapons Convention) has an annual budget of just \$1.4 million—less than the average McDonald’s restaurant.⁵⁴ The entire spending on reducing

existential risks from advanced artificial intelligence is in the tens of millions of dollars, compared with the billions spent on improving artificial intelligence capabilities.⁵⁵ While it is difficult to precisely measure global spending on existential risk, we can state with confidence that humanity spends more on ice cream every year than on ensuring that the technologies we develop do not destroy us.⁵⁶

In scientific research, the story is similar. While substantial research is undertaken on the risk of smaller catastrophes, those that could destroy humanity's longterm potential are neglected. Since 1991 there have been only two published climate models on the effects of a full-scale nuclear war between the United States and Russia, even while hundreds of missiles remain minutes away from a possible launch.⁵⁷ There has been tremendous work on understanding climate change, but the worst-case scenarios—such as those involving more than six degrees of warming—have received comparatively little study and are mostly ignored in official reports and policy discussions.⁵⁸

Given the reality and importance of existential risks, why don't they already receive the attention they need? Why are they systematically neglected? Answers can be found in the economics, politics, psychology and history of existential risk.

Economic theory tells us that existential risk will be undervalued by markets, nations and even entire generations. While markets do a great job of supplying many kinds of goods and services, there are some kinds that they systematically undersupply. Consider clean air. When air quality is improved, the benefit doesn't go to a particular individual, but is shared by everyone in the community. And when I benefit from cleaner air, that doesn't diminish the benefit you get from it. Things with these two properties are called *public goods* and markets have trouble supplying them.⁵⁹ We typically resolve this at a local or national level by having governments fund or regulate the provision of public goods.

Protection from existential risk is a public good: protection would benefit us all and my protection doesn't come at the expense of yours. So we'd expect existential risk to be neglected by the market. But worse, protection from existential risk is a *global public good*—one where the pool of beneficiaries spans the globe. This means that even nation states will neglect it.

I am writing this book in the United Kingdom. Its population of nearly 70 million ranks it as one of the more populous countries in the world, but it contains less than 1 percent of all the people alive today. If it acts alone on an existential risk, it bears the full cost of the policy, yet only reaps a hundredth of the benefit. In other words, even if it had a well-informed government acting in the longterm interests of its citizens, it would undervalue work on existential risk by a factor of 100. Similarly, Russia would undervalue it by a factor of 50, the United States by a factor of 20, and even China would undervalue it by a factor of five. Since such a large proportion of the benefits spill out to other countries, each nation is tempted to free-ride on the efforts of others, and some of the work that would benefit us all won't get done.

The same effect that causes this undersupply of protection causes an oversupply of risk. Since only 1 percent of the damages of existential catastrophe are borne by the people of the United Kingdom, their government is incentivized to neglect the downsides of risk-inducing policies by this same factor of 100. (The situation is even worse if individuals or small groups become able to pose existential risks.)

This means management of existential risk is best done at the global level. But the absence of effective global institutions for doing so makes it extremely difficult, slowing the world's reaction time and increasing the chance that hold-out countries derail the entire process.

And even if we could overcome these differences and bargain toward effective treaties and policies on existential risk, we would face a final problem. The beneficiaries are not merely global, but intergenerational—all the people who would ever live. Protection from existential risk is an *intergenerational global public good*. So even the entire population of the globe acting in concert could be expected to undervalue existential risks by a very large factor, leaving them greatly neglected.⁶⁰

Additional reasons can be found in political science. The attention of politicians and civil servants is frequently focused on the short term.⁶¹ Their timescales for thought and action are increasingly set by the election cycle and the news cycle. It is very difficult for them to turn their attention to issues where action is required now to avert a problem that won't strike for several election cycles. They are unlikely to get punished for letting it slide and many more urgent things are clamoring for attention.

One exception to this is when there is an active constituency pushing for the early action: their goodwill acts as a kind of immediate benefit. Such constituencies are most powerful when the benefits of the policy are concentrated among a small fraction of society, as this makes it worth their while to take political action. But in the case of existential risk, the benefits of protection are diffused across all citizens, leaving no key constituency to take ownership of the issue. This is a reason for neglect, albeit one that is surmountable. When citizens are empathetic and altruistic, identifying with the plight of others—as we have seen for the environment, animal rights and the abolition of slavery—they can be enlivened with the passion and determination needed to hold their leaders to account.

Another political reason concerns the sheer gravity of the issue. When I have raised the topic of existential risk with senior politicians and civil servants, I have encountered a common reaction: genuine deep concern paired with a feeling that addressing the greatest risks facing humanity was “above my pay grade.” We look to our governments to manage issues that run beyond the scope of our individual lives, but this one runs beyond the scope of nations too. For political (as well as economic) reasons it feels like an issue for grand international action. But since the international institutions are so weak, it is left hanging.

Behavioral psychology has identified two more reasons why we neglect existential risk, rooted in the heuristics and biases we use as shortcuts for making decisions in a complex world.⁶² The first of these is the *availability heuristic*. This is a tendency for people to estimate the likelihood of events based on their ability to recall examples. This stirs strong feelings about avoiding repeats of recent tragedies (especially those that are vivid or widely reported). But it means we often underweight events which are rare enough that they haven’t occurred in our lifetimes, or which are without precedent. Even when experts estimate a significant probability for an unprecedented event, we have great difficulty believing it until we see it.

For many risks, the availability heuristic is a decent guide, allowing us to build up methods for managing the risk through trial and error. But with existential risks it fails completely. For by their very nature, we never have any experience of existential catastrophe before it is too late. If only seeing is believing, we will step blindly over the precipice.

Our need for vividness also governs our altruistic impulses. As a society,

we are good at *acute* compassion for those in peril—for the victims of a disaster we can see in the news reports. We may not always act, but we certainly feel it. We sit up, our hearts in our throats: fearing for their safety, mourning for their loss. But what we require is a more expansive compassion; a more imaginative compassion; one that acts over the long term, recognizing the humanity of people in distant times as well as distant places.

We also suffer from a bias known as *scope neglect*. This is a lack of sensitivity to the scale of a benefit or harm. We have trouble caring ten times more about something when it is ten times as important. And once the stakes get to a certain point, our concern can saturate.⁶³ For example, we tend to treat nuclear war as an utter disaster, so we fail to distinguish nuclear wars between nations with a handful of nuclear weapons (in which millions would die) from a nuclear confrontation with thousands of nuclear weapons (in which a thousand times as many people would die, and our entire future may be destroyed). Since existential risk derives its key moral importance from the size of what is at stake, scope neglect leads us to seriously underweight its importance.

These reasons for the neglect of existential risk present a formidable challenge to it ever receiving due concern. And yet I am hopeful. Because there is a final reason: existential risk is very new. So there hasn't yet been time for us to incorporate it into our civic and moral traditions. But the signs are good that this could change.

Humans must have contemplated the end of humanity from the earliest times. When an isolated band or tribe died out during a time of extreme hardship, the last survivors will have sometimes wondered whether they were the last of their kind, or whether others like them lived on elsewhere. But there appears to have been very little careful thought about the possibility and importance of human extinction until very recently.⁶⁴

It wasn't until the mid-twentieth century, with the creation of nuclear weapons, that human extinction moved from a remote possibility (or a certainty remote in time) to an imminent danger. Just three days after the devastation of Hiroshima, Bertrand Russell began writing his first essay on the implications for the future of humanity.⁶⁵ And not long after, many of the scientists who created these weapons formed the *Bulletin of the Atomic Scientists* to lead the conversation about how to prevent global

destruction.⁶⁶ Albert Einstein soon became a leading voice and his final public act was to sign a Manifesto with Russell arguing against nuclear war on the explicit grounds that it could spell the end for humanity.⁶⁷ Cold War leaders, such as Eisenhower, Kennedy and Brezhnev, became aware of the possibility of extinction and some of its implications.⁶⁸

The early 1980s saw a new wave of thought, with Jonathan Schell, Carl Sagan and Derek Parfit making great progress in understanding what is at stake—all three realizing that the loss of uncounted future generations may overshadow the immediate consequences.⁶⁹ The discovery that atomic weapons may trigger a nuclear winter influenced both Ronald Reagan and Mikhail Gorbachev to reduce their country's arms and avoid war.⁷⁰

And the public reacted too. In 1982, New York's Central Park saw a million people come together to march against nuclear weapons. It was the biggest protest in their nation's history.⁷¹ Even in my birthplace of Australia, which has no nuclear weapons, we joined the global protest—my parents taking me with them on marches when I was just a small child they were fighting to protect.

In this way, existential risk was a highly influential idea of the twentieth century. But because there was one dominant risk, it all happened under the banner of nuclear war, with philosophers discussing the profound new issues raised by “nuclear ethics,” rather than by “existential risk.” And with the end of the Cold War, this risk diminished and the conversation faded. But this history shows that existential risk is capable of rousing major global concern, from the elite to the grass roots.

Modern thinking on existential risk can be traced through John Leslie, whose 1996 book *The End of the World* broadened the focus from nuclear war to human extinction in general. After reading Leslie's work, Nick Bostrom took this a step further: identifying and analyzing the broader class of existential risks that are the focus of this book.

Our moral and political traditions have been built over thousands of years. Their focus is thus mostly on timeless issues that have been with us throughout our history. It takes time to incorporate the new possibilities that our age opens up, even when those possibilities are of immense moral significance. Existential risk still seems new and strange, but I am hopeful that it will soon find its way into our common moral traditions. Environmentalism burst in upon the global political scene less than twenty

years before I was born, and yet I was raised in a milieu where it was one of the main parts of our moral education; where the earlier disregard for the environment had become unthinkable to my generation. This can happen again.

One of my principal aims in writing this book is to end our neglect of existential risk—to establish the pivotal importance of safeguarding humanity, and to place this among the pantheon of causes to which the world devotes substantial attention and resources. Exactly how substantial remains an open question, but it clearly deserves far more focus than it has received so far. I suggest we start by spending more on protecting our future than we do on ice cream, and decide where to go from there.

We have now seen the broad sweep of human history, the size of humanity's potential, and why safeguarding our future is of the utmost importance. But so far, you have mostly had to take my word that we do face real risks. So let us turn our attention to these risks, examining the key scientific evidence behind them and sorting out which ones are most worthy of our concern. The next three chapters explore the natural risks we have faced throughout our history; the dawn of anthropogenic risks in the twentieth century; and the new risks we will face over the century to come.