

# Securing posterity

Written by Leopold Aschenbrenner, published on Oct. 19, 2020

Ever since the development of the atomic bomb, humanity has possessed the means for its own destruction. There have, of course, always been natural risks of human extinction: an asteroid impact or a supervolcano eruption could make the planet uninhabitable. But since that fateful first nuclear test in the New Mexico desert, the [primary threat](#) to the survival of human civilization has come from man himself.

In the decades since, the risks from human activity have multiplied. While the median climate change scenario would merely be unpleasant, there is the possibility of [tail-end scenarios](#) that could end civilization as we know it. Advances in biotechnology might allow us to engineer a new pathogen that silently spreads before killing not just millions, but all of humanity. Some, like Elon Musk, fear that we might soon develop artificial intelligence that could go rogue.

Philosophers like [Nick Bostrom](#), [Derek Parfit](#), and [Toby Ord](#) have become increasingly concerned about such so-called “existential risks.” An unrecoverable collapse of civilization wouldn’t just be tragic for the billions who would suffer and die. Perhaps the greatest tragedy would be the foreclosing of all of humanity’s potential. Humanity could flourish for billions of years and enable trillions of happy human lives—if only we do not destroy ourselves beforehand.

This line of thinking has led some to question whether “progress”—in particular, technological progress—is as straightforwardly beneficial as commonly assumed. Nick Bostrom imagines the process of technological development as “[pulling balls out of a giant urn](#).” So far, we’ve been lucky, pulling out a great many “white” balls that are broadly beneficial. But someday, we might pull out a “black” ball: a new technology that destroys humanity. Before that first nuclear test, some of the physicists worried that the nuclear bomb would ignite the atmosphere and [end the world](#). Their calculations ultimately deemed it “extremely unlikely,” and so they proceeded with the test—which, as it turns out, did not end the world. Perhaps the next time, we don’t get so lucky.

The same technological progress that creates these risks is also what drives economic growth. Does that mean economic growth is inherently risky? Economic growth has brought about extraordinary prosperity. But for the sake of posterity, must we choose safe stagnation instead? This view is arguably becoming ever-more popular, particularly amongst those concerned about climate change; Greta Thunberg recently denounced “[fairy tales of eternal economic growth](#)” at the United Nations.

I argue that the opposite is the case. *It is not safe stagnation and risky growth that we must choose between; rather, it is stagnation that is risky and it is growth that leads to safety.*

We might indeed be in “time of perils”: we might be advanced enough to have developed the means for our destruction, but not advanced enough to care sufficiently about safety. But stagnation does not solve the problem: we would simply stagnate at this high level of risk. Eventually, a nuclear war or environmental catastrophe would doom humanity regardless.

Faster economic growth could initially increase risk, as feared. But it will also help us get past this time of perils more quickly. When people are poor, they can’t focus on much beyond ensuring their own livelihoods. But as people grow richer, they start caring more about things like the environment and protecting against risks to life. And so, as economic growth makes people richer, they will invest more in safety, protecting against existential catastrophes. As technological innovation and our growing wealth has allowed us to conquer past threats to human life like smallpox, so can faster economic growth, in the long run, increase the overall chances of humanity’s survival.

This argument is based on a [recent paper](#) of mine, in which I use the tools of economic theory—in particular, the standard models economists use to analyze economic growth—to examine the interaction between economic growth and the risks engendered by human activity.

In this model, society must choose how much of its resources to allocate to consumption and how much to safety efforts. Consumption makes us happy, but also creates risks of catastrophe. Investing in safety can in turn help mitigate that risk.

For example, consuming fossil fuels can engender great prosperity, but also increases the risk of tail-end climate change. We can spend money on carbon abatement to reduce this risk. Or

consider air travel. It's very useful as well, but also facilitates the spread of infectious diseases, including potentially a pandemic that could wipe out the human race. We can spend money on pandemic preparedness to mitigate that risk.

Crucially, society is impatient; it discounts the future. People generally care most about their more immediate well-being. Although they may care about their kids and grandkids, they are certainly not particularly concerned about the trillions of potential lives billions of years in the future that the aforementioned philosophers appeal to.

However, an impatient society *does* care about not getting wiped out. Therefore, what fraction of its resources this impatient society will allocate to safety depends on how much the people in this society value their own lives.

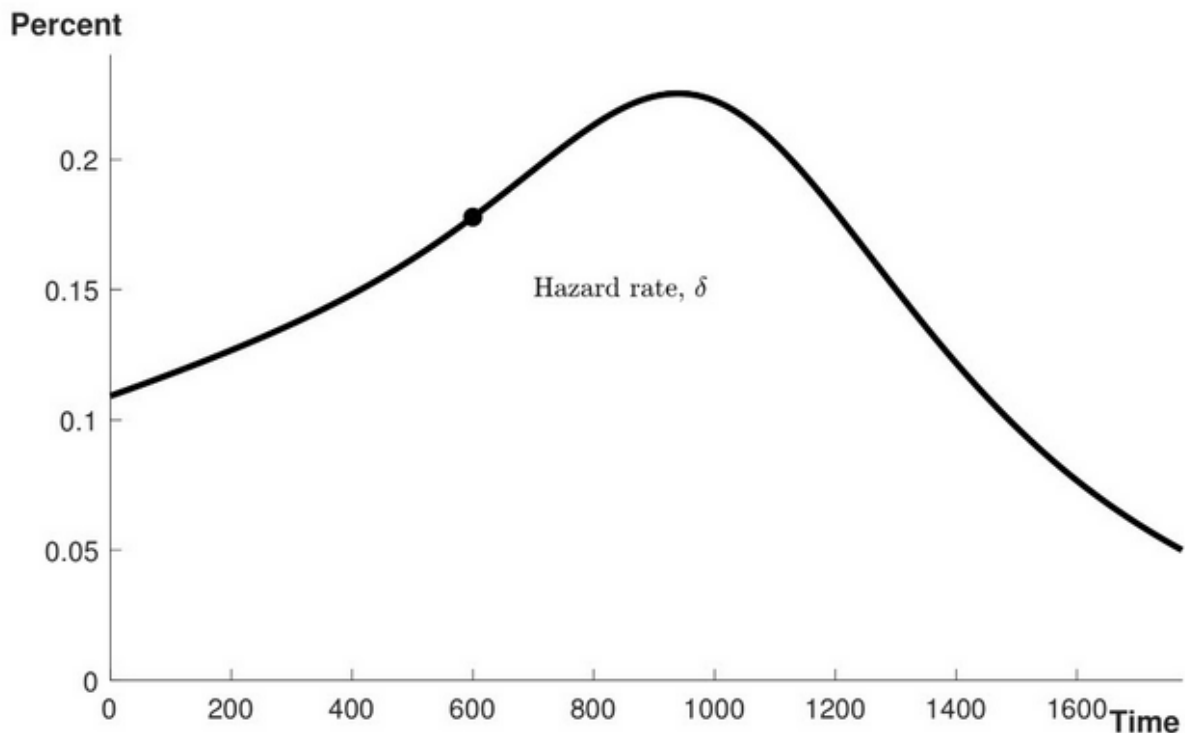
As it turns out, under the standard preferences used in economic theory, people value life more and more as they grow richer. This is because of the diminishing marginal returns of consumption. As you grow richer, using an extra dollar to purchase more consumption goods gives you less and less additional utility; meanwhile, as your life becomes better and better, you stand to lose more and more if you die. As a result, the richer people are, the greater the fraction of their income they are willing to sacrifice to protect their lives.

Comparing the current pandemic to the 1918 pandemic illustrates this phenomenon. Today, we are putting much of life on hold to minimize deaths. By contrast, in 1918, nonpharmaceutical interventions were milder and went on only [for a month on average](#) in the U.S., even though the Spanish Flu was arguably deadlier and claimed younger victims. We are willing to sacrifice much more today than a hundred years ago to prevent deaths because we are richer and thus value life much more.

What does this mean for our model? Initially, a poor society will start out by allocating nearly all of its resources to consumption. And so as the economy grows, so does risk.

However, as people grow richer, they start valuing life more. They start investing in safety to mitigate risk, shifting more and more resources from consumption to safety. At this point, as the economy grows, risk begins to fall.

The risk of an existential catastrophe then looks like an inverted U-shape over time:



The dot represents where we might be right now. Over the past centuries, as we have grown out of poverty, we have overwhelmingly focused on consumption. As a result, risk is growing.

But as we are growing richer, we are beginning to value life more, and are slowly investing more in safety. Eventually, we will have shifted enough resources to safety such that risk begins to fall—fall exponentially to zero, in fact, such that there is a positive probability of humanity surviving to reach a grand future. And all of this occurs despite our society’s impatience.

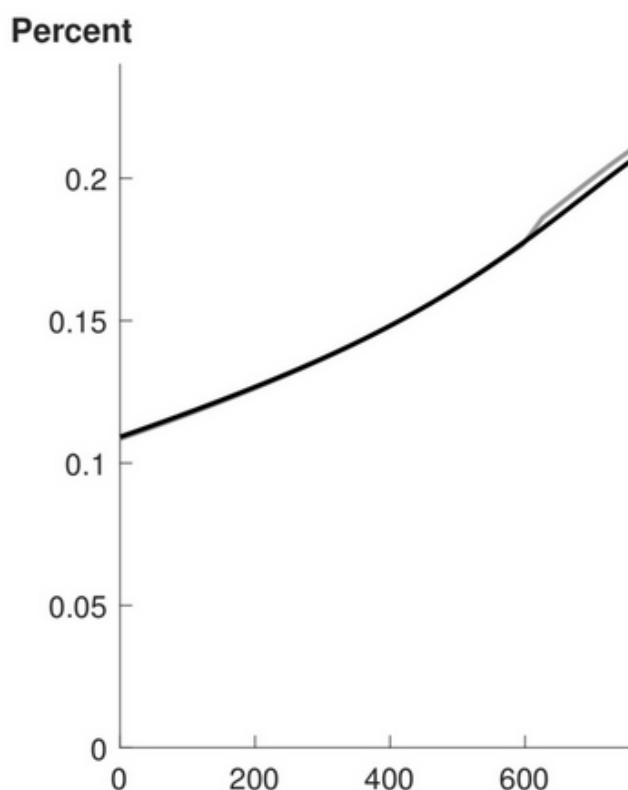
There is an analog to this in environmental economics, called the “[environmental Kuznets curve](#).” It was theorized that pollution initially rises as countries develop, but, as people grow richer and begin to value a clean environment more, they will work to reduce pollution again. That theory has arguably been vindicated by the path that Western countries have taken with regard to water and air pollution, for example, over the past century.

The idea that we are in a unique time in history in which we are facing an elevated risk of existential catastrophe is not new either. Carl Sagan was the one who coined the term “time of

perils.” Derek Parfit called it the “hinge of history.” They [argue](#) that the discoveries of the last centuries have granted humanity immense power, and so we are in a most “dangerous and decisive” period. But if we manage to survive, our descendants will be able to spread throughout the galaxy, making us much less vulnerable. They will have mastered new technologies that make us immune to bioengineered pathogens, neutralize the threat from atomic bombs, provide plentiful energy without destroying the environment, and keep artificial intelligence in check so it faithfully serves human needs. With their technology and wisdom, our descendants will be able to secure a long and safe future. Our challenge, then, is to make it through this unique perilous period.

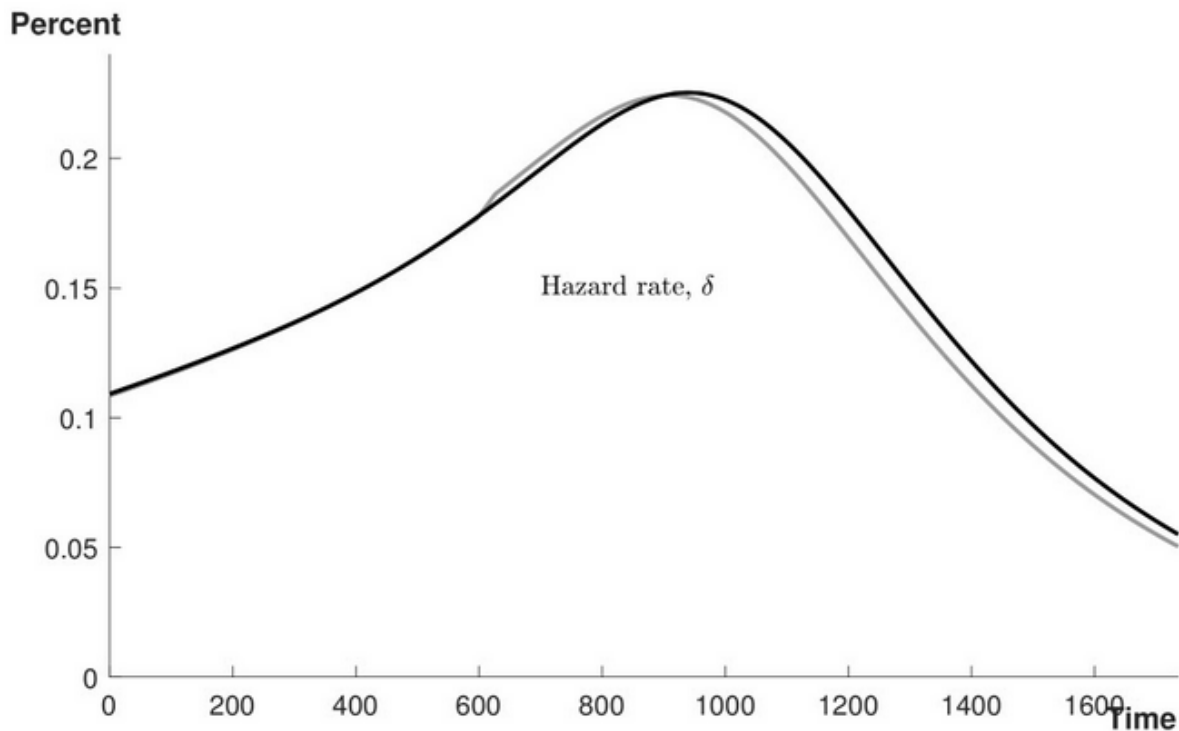
Seeing the rising levels of existential risk over the past centuries, some might call for an end to economic growth. They might argue, rightfully so, that economic growth has only led to rising risk in the past.

Indeed, a period of accelerated economic growth would initially also accelerate the rise in risk. The level of risk might look something like this, where the lighter line is the path with accelerated growth:



Even a few hundred years later, the critics of growth would seem to be vindicated! Faster growth just increased the risk!

Except that they are missing the whole picture:



The accelerated economic growth also accelerated our path along the inverted-U shape of risk. Faster growth means people are richer sooner, so they value life more sooner, so society shifts resources to safety sooner—and ultimately we will begin the decline in risk sooner. As a result, the overall probability of an existential catastrophe—the area under the risk curve—declines!

Faster growth means we get through the “time of perils” more quickly. Indeed, stagnation would be the most dangerous choice of all: we would be stuck at an elevated level of risk, meaning an eventual existential catastrophe would be inevitable.

It is as though we are on a treacherous voyage to cross a stormy ocean. If we stay put, it is only a matter of time before a powerful enough gust of wind or large enough wave capsizes our ship. Rather, we should sail with haste, spending as little time as possible at the mercy of the dangerous seas before we reach the safe shores.

Now, there is another critical parameter: how effective are our safety efforts at mitigating risks, compared to how much our consumption activity increases them? We might think of this as the “fragility” of humanity.

I have been focusing on the case in which humanity is moderately fragile. That’s because this is the scenario in which our future is in doubt.

There are two other scenarios. On the one extreme, humanity is not very fragile. Safety efforts easily mitigate any risk. Perhaps just a little bit of pandemic preparedness abates the risk of air travel; perhaps just a [few people](#) working on AI safety can ensure that AI doesn’t go rogue. In this sort of world, we would never see an initial rise in risk, which wouldn’t accord with the experiences of the past centuries. Moreover, in this scenario, economic growth naturally decreases risk—so even if it turns out that we do live in this world, minimizing risk would demand growing as quickly as possible.

On the other extreme, humanity is extremely fragile. No matter how high a fraction of our resources we dedicate to safety, we cannot prevent an unrecoverable catastrophe. Perhaps weapons of mass destruction are simply too easy to build, and no amount of even totalitarian safety efforts can prevent some lunatic from eventually causing nuclear annihilation. We indeed might indeed be living in this world; this would be the model’s version of Bostrom’s “[vulnerable world hypothesis](#),” Hanson’s “[Great Filter](#),” or the “[Doomsday Argument](#).” Moreover, risk would naturally increase with growth in this world, so much so that even an impatient society might choose stagnation. However, from the point of view of posterity, there is nothing we can do regardless. An existential catastrophe is inevitable, and it is impossible for us to survive to reach a grand future. So even if there is some probability we do live in this world, to maximize the moral value of the future, we should act as if we live in the other scenarios where a long and flourishing future is possible.

To be clear, this is just one imperfect theoretical model. For example, the model considers the optimal allocation. But safety investment to protect against existential catastrophe is a global public good—and as the example of climate change illustrates, the provision of global public goods can be challenging. Or we might model the source of risk differently, e.g. as coming directly from the technological development process itself rather than from the human activity the technologies are used for.

In that sense, you shouldn't take this model as any sort of definitive answer. But, indeed, a definitive answer is not the point of economic theory. What all the mathematical machinery of economic theory [can do](#) is reveal intuitive insights that we might have otherwise not considered. And the intuitive story—that we want to get through the time of perils as quickly as possible—seems eminently plausible.

The model also suggests a broader insight. Making people richer doesn't just improve their well-being, but it can also change what they value. In this case, people value life more as they grow richer, and valuing life more leads them to care more about reducing existential risk.

In that sense, making people richer—whether through growth or other means—can accomplish similar goals as trying to persuade people to adopt different values. For example, some in the “Effective Altruism” community have pursued an intellectual project called “[longtermism](#)” to get people to care more about the very long run—in effect, to get them to lower their discount rate. In terms of getting people to care more about preventing existential risk, I calculate that doubling people's consumption might have a similar effect as, say, lowering people's discount rate from 2% to 1.4%. Perhaps, if we followed this argument to the end, we might reach the counterintuitive conclusion that the most effective thing we can do reduce the risk of an existential catastrophe is not to invest in safety directly or to try to persuade people to be more long-term oriented—but rather to spend money on alleviating poverty, so more people are well-off enough to care about safety.

The model also suggests a potential downside to some of the well-meaning efforts that aim to directly increase people's concern about catastrophes. An impatient society might in fact favor slower growth for the sake of safety. Recall how faster growth initially increased risk in the short term, and how it took several hundred years before we saw an overall reduction in risk. If people are mostly concerned about more immediate risks to themselves, they might deliberately decide to slow growth, reducing the risk for their current generation at the cost of greater risk for future generations.

Of course, that is not what those concerned about existential catastrophes are advocating. But we should worry about the culture they would further if their ideas about the risks of technological progress continue to become more mainstream. We may end up with ever-more safety bureaucracy and ever-more risk aversion, not contributing much to decreasing existential



risk directly but hindering innovation and slowing growth—ultimately reducing the chances of humanity’s survival.

Arguably, that’s exactly what we have been doing in the United States. In the spirit of “an abundance of caution,” we’ve been overzealously applying the [precautionary principle](#). This has already meant we are less able to deal with catastrophes. The regulatory bureaucracy at the FDA meant the U.S. was unable to detect the coronavirus hitting our shores in February and is delaying the development of effective therapeutics and vaccines. Proponents claim the precautionary principle puts safety first, but, in the long run, it may actually make us less safe.

To the extent that we are at a crossroads between continued growth with—potentially risky—innovation on the one hand, and stagnant decadence in the name of comfort and safety on the other hand, those concerned for posterity should respond with a unified voice: we choose growth.

Aschenbrenner, Leopold (2020): Securing posterity, Online Article 19.10.2020, Issue 2: <https://worksinprogress.co/issue/securing-posterity/>, last access 22.10.2020.