The Summit of the Future is a Chance to 'Human-Proof' Our Global Governance System

Adam Day

COP28 was a reminder that vested interests and powerful actors can prevent the kind of bold, transformative action we need to stop catastrophic climate change. But, climate change is not the only risk we face today. Toby Ord, a leading thinker on global risk, estimates that a humanity-ending catastrophe in the next one hundred years has a one in six chance of happening.[1] If he is even remotely close to the mark, we are gambling humanity on the roll of a die, or a game of Russian roulette. It may even be closer to a coin flip if you believe other estimates.[2] Some experts have warned that we may be sliding into a range of dystopian futures, where a global surveillance authority controls much of our lives, a race of humans enhanced by artificial intelligence (AI) rises to rule us all, or global warming rises by an unbearable 8°C.[3] These may sound like far-fetched scenarios, but the reality of planetary-level threats should be fresh in our minds after our COVID-19 experience. The 13,000 nuclear weapons currently on a hair-trigger amidst deepening geopolitical crises should have us worried as a species.

The Summit of the Future in 2024 has been described as a forum to 'future proof' global governance, building the necessary foresight, emergency response, and future-oriented institutions needed to address what the Secretary-General has called today's "polycrisis."[4] The UN General Assembly resolution setting out the scope of the Summit has an entire track focused on future generations, and another on improving global governance. I want to make the case in this brief that future-proofing global governance also means 'human proofing' it by thinking in systemic terms about long-term, catastrophic risks. We humans have evolved a set of serious shortcomings, myopias, and deeply-entrenched path dependencies that make us very badly placed to manage global catastrophic risks. The Summit of the Future is a unique moment to address these head on, get our noses off the grindstone of crisis management, and begin to design an architecture that will shape our collective behaviour for the coming storms.

Why We Undervalue Massive Future Risks

We are terrible at thinking about the future. Five related human shortcomings tend to make us undervalue future risks and focus far too much on immediate crisis response.

1. Humans display a **'presentist bias'** that means we consistently favour the present over the future.[5] Democracies tend to have the worst cases of presentism, as electoral cycles and public opinion have a direct impact on political decisions.[6] But in general, people tend to give far greater weight to imminent risks than more distant ones, even if the distant ones are much greater.

2. We fall back on an **'availability heuristic'** where we tend to estimate risk based on our own experiences. Yes, we might have an acute sense of the risks of future pandemics with COVID-19 fresh in our minds, but most of the global catastrophic risks have no precedent in our lifetimes. We have never experienced 8°C of global warming; we don't know what a nuclear war would feel like; we have never lived under the thumb of AI. And when it comes to

existential threats, these have by definition never occurred.[7] Massive, unprecedented harms tend to be chronically undervalued.[8]

3. We suffer from **'scope neglect'** in which we cannot absorb the reality of massive harm to large numbers of people.[9] We struggle to care one thousand times more about something that is one thousand times more harmful. Worse, the suffering of a small number of people often feels more real and vivid than the suffering of millions. This is sometimes called psychological 'numbing' or 'moral disengagement' from the large-scale impacts of our own actions.[10] It is almost unimaginable to envision a global nuclear war resulting in the death of everyone, regardless of how likely it may be.

4. We tend to set **unrealistic scientific standards** for managing global catastrophic risks. Most big global policy change is developed after decades of scientific research, leading to a sufficiently large consensus that can overcome vested interests. Look at the overwhelming preponderance of evidence linking smoking to cancer, and even today smoking is legal. After 40 years of rigorous scientific findings that human activity is causing climate change, COP28 still couldn't reach consensus on phasing out fossil fuels. Monday Night Football is a testament to our willingness to continue a sport that clearly causes depression and suicide due to repeated concussions. But policy lags behind science. We cannot afford to build this kind of cumulative scientific consensus over decades when it comes to existential and catastrophic risks. Carl Sagan wisely pointed out: "Theories that involve the end of the world are not amenable to experimental verification – or at least, not more than once."[11] We don't have time to get 100 per cent of scientists to agree on the dangers of solar radiation management, or AI, or the next deadly pathogen.[12]

5. We suffer from **crisis-driven attention deficit disorder**. Quick, let's list some catastrophic risks: climate change, collapsing ecosystems, mass species extinction, runaway plastic pollution, food system collapse, synthetic pathogens, US-China tensions over Taiwan, Russia's nuclear saber-rattling on Ukraine, potential spillover from the Gaza conflict, soaring inequality, cyber-vulnerabilities in our critical infrastructure, and on and on and on. Surrounded by crises, we tend to rush from one to the next, like a firefighter with a single hose and multiple fronts. We forget the last crisis as all our attention and resources go to the next one (we are seeing this today as the world seems to have forgotten the war in Ethiopia, is forgetting the war in Ukraine, and frontlines are all about Gaza, for now). This also means we fail to balance short- and long-term risks.[13] Will AI transform our lives for the better or destroy us all? Without certainty to this question, we fall into policy whiplash, rushing from overregulation to laissez faire market approaches. And when a crisis does happen – like the 2008 financial collapse – we have a strong tendency to try to get back to the status quo ante as the safest bet. But, as the 2008 crisis shows, our willingness to shore up the existing system has baked in some of the worst aspects of our global financial architecture, contributing to soaring inequality and a market that is almost sure to collapse again.

The combined result of these human shortcomings is that the global governance system has become a Rube Goldberg machine that generates endless acts of coordination and small-scale responses to the most immediate crisis but fails to drive a deeper transformation.[14] Global governance reform is reduced to an incremental series of fixes that solve immediate problems, but also creates new layers of bureaucracy and coordination. We seem destined for a "tragedy of the uncommons"[15] where our systems are, by design, unable to rise to the challenge of systemic risk.[16]

Maybe all of this should not be such a surprise. In 1945 when the atomic bomb was first tested, the Manhattan Project scientists genuinely thought there was a chance it would ignite the atmosphere and destroy all life on the planet. Knowing this, they went ahead with the detonation. They were wrong, but that does not negate the fact that America's leaders were willing to risk humanity in the name of getting the bomb before Hitler. Today we aren't much different. We elect leaders who promise to maintain stockpiles of thousands of nuclear weapons, continue to subsidize fossil fuels, invest in untested geoengineering experiments, accelerate the development of AI beyond our ability to control its safety, and develop bioweapons that could end all life on this planet.

Our common human intellectual and moral poverty about these risks leads to a question: How should we make decisions about the future?[17] How can we design a global governance architecture that can more accurately reflect and manage these risks? I have reached the conclusion that this will require us to 'human-proof' global governance, building designs that combat our evolutionary shortcomings. The first step is to think systematically about the future.

Thinking About the Future

Imagine a person living far away from you on the other side of the world, or even on the space station. I hope we can agree that this person's life is no less valuable than our own. I hope we can agree that human life has equal value regardless of where it is. But people's lives are also equally valuable regardless of *when* they are.[18] A human born tomorrow is no less valuable than one alive today. Imagine yourself ten years ago: you were no less human then than you are today, and you will be just as human tomorrow.

But we actively discriminate against future generations all the time. When we burn fossil fuels, or generate nuclear waste, or put plastics into the ocean, or add to the debt, we are taking actions that have a negative impact on future generations. Yet we have no accountability for them.[19] When governments make policy, they are only accountable to voters (not children or future generations). Government policy responds to our presentist bias and also favours the present. [20] As a result, we have generated a system based on legalized and normalized theft from the future, where our global governance systems have clear lines of accountability towards living generations, but near absolute impunity towards the future.

Barring the end of the world, we know that lots of people will be born and we know that they will need many of the same things we need today.[21] But how should we account for those needs, and how should we balance them against the very real needs of people today? There are a lot of ethical rabbit holes and interesting philosophical arguments here, and I encourage you to explore the resources footnoted here on questions of long-termism, effective altruism, and the thorny issues that come into play when we try to balance present and future generations.[22]

But we don't need to get too far into the weeds. The future may involve trillions of people, and they may inhabit distant parts of the galaxy.[23] But if we focus too much on this potentially huge future population, long-term considerations could 'swamp' short-term ones.[24] Let's instead stick with an uncontroversial argument: ending humanity would be a very bad thing, and making the lives of future generations miserable or unlivable would also be very bad.[25]

If ending humanity is a global bad, then we should think of the global governance of existential risks as a global public good. Everyone benefits from our collective survival. Like clean air, a global public good is a collective, shared benefit.[26] But this does not mean that everyone is equally incentivized to invest in it. Why should one country pay for a global public good when its citizens will only see a fraction of a benefit that can be delivered by others? Why not just wait for others to invest and then

reap the rewards of a globally-shared good? This is the free-rider problem. Future generations would clearly benefit from a decision to phase out fossil fuels now, but that does not mean all countries are equally incentivized to pay for the transition to green energy.[27]

It is easy to say "we need collective action" in the face of global risks to future generations. But it is just as easy to say that collective action is extremely unlikely in today's world. The fact that we are custodians of a planet that will be passed to future generations does not mean we will act as fiduciaries for those people.[28] In fact, I think the COP28 process should be a stark reminder that when push comes to shove, our reliance on collective action is just not good enough.

Big Ideas for Big Risks

Faced with these global catastrophic risks, we should be able to come up with some equally big, bold ideas. There is some good news: a growing interest in studying catastrophic risks has generated new initiatives and dedicated centres of thinking.[29] These have produced an array of innovative proposals, including quite a lot that draws on systems thinking.[30] The less good news: most of the boldest proposals remain mired in academic centres, not on the desks of today's leaders. Instead, I see the main proposals on the table today falling into three broad categories: (1) a world government or centralized body to address catastrophic risks; (2) a science-policy interface that can help us understand future risks; or (3) a focal point for the future. These are interesting but may need some additional thought to make our global governance system fit for the future.

(1) A World Government?

In 1948, Albert Einstein saw the destructive force of nuclear weapons and called for a "world government." In his view, this was necessary because "there is no other possible way of eliminating the most terrible danger in which man has ever found himself."[31] Of course, no serious politician today would dare to utter the phrase 'world government,' and the idea that a single actor could regulate global affairs is completely off the table. As Anne-Marie Slaughter noted in her landmark *A New World Order*: "a world government is both infeasible and undesirable."[32]

But in some arenas the idea of a centralized governance body for global catastrophic risks may be making a comeback. Nick Bostrom has described a scenario where a super-intelligent AI "singleton" could manage global existential risks.[33] Similarly, environmental experts have long called for some version of a 'Global Environmental Council,' a 'World Environmental Court,' or a 'World Environmental Organization.'[34] Other ideas include repurposing the UN Trusteeship Council to represent future generations, or creating a 'Global Resilience Council' to address systemic threats to humanity.

(2) An IPCC for Global Catastrophic Risks

During the consultations I've attended in the past two years, some version of "we need an IPCC for existential risks" was a very common phrase. In fact, a science-policy interface for planetary-level risks has been put forward by many of the most important expert organizations, including:

- A high-level UN panel supported by a scientific commission to evaluate risks and propose actions in line with a UN 'global action plan' for systemic crises.[35]

- A science-policy interface for catastrophic and existential risks.[36]

- An IPCC model of scientific inquiry into existential risk.[37]

- An International Panel on Global Catastrophic Risk.[38]

- An Emergency Platform convened by the UN Secretary-General to deal with global risks as they are unfolding.[39]

These all share a common theory of change: if we can get the science to connect to the policy, we can orient global governance around systemic risks.

(3) A Minister for the Future

The Secretary-General is planning to appoint an Envoy for Future Generations, and he has already formed the UN Futures Lab Network to expand the UN's foresight capacities.[40] Fans of Kim Stanley Robinson may catch the apparent reference to *The Ministry for the Future*.[41] But the proposal for a special envoy comes most directly from national envoys in places like Wales and New Zealand, which have demonstrated real success.[42] Their experiences offer many good ideas about future-proofing governments that should be explored.[43]

These three clusters of proposals could make a big difference, but they are also worrying because they look a lot like what has been tried in the past. To quote Oran Young's critique of today's global governance, they seem more like "familiar recipes" and "formulaic prescriptions" that may be easy to grasp but have an almost unbroken track record of failure in the past.[44] Some of them are also clearly outside the window of political possibility anytime soon. Even the creation of an Emergency Platform (put forward by the Secretary-General last year) was seen as contentious, despite being quite feasible. If Member States are unable to agree on that, are we really expecting them to set up a world government or repurpose the Trusteeship Council for future generations? It seems unlikely. We need creative, doable proposals for the Summit of the Future and beyond.

## Four Ways Systems Thinking Can Help Us Human-Proof Global Governance

Systems thinking can help us frame future risks in a more balanced way and offers some innovative ideas for the Summit of the Future. Here are four arguments that systems thinking (1) moves us beyond thinking of resilience as just a recovery from a short-term shock; (2) demands a diversity of decision-makers; (3) enables us to put 'brakes' on risks when they reach tipping points; and (4) pushes us towards transformative, not reactive governance. This then leads into some concrete ideas for the Summit of the Future.

(1) Beyond Resilience

Systems thinking exposes a simple truth about the world: harm is non-linear. Of course, we tend to think of harm as cause and effect. We put a chemical into a stream, someone downstream is harmed. We detonate a nuclear weapon, millions die. But in our planetary system, harm can be multidirectional, indirect, and change over time.[45] Some actions can start small and snowball – biodiversity loss and

melting Arctic ice shelfs, for example. Some harms are big enough to close off possibilities for the future. A pandemic that eradicated 90 per cent of the world would close off a huge number of potential scenarios for 2045. Some harms we should think as systemic, changing the overall range of possibilities in the future. The Industrial Revolution, for example, caused a fundamental shift in our means of production, but it also locked in certain path dependencies to unsustainable energy consumption.

Systems have "points of peak importance," where the most significant benefits and costs tend to converge.[46] The peak may look like a crisis, where the range of possible outcomes suddenly expands

Potential for human well-being

Global catastrophe

The next 100 years
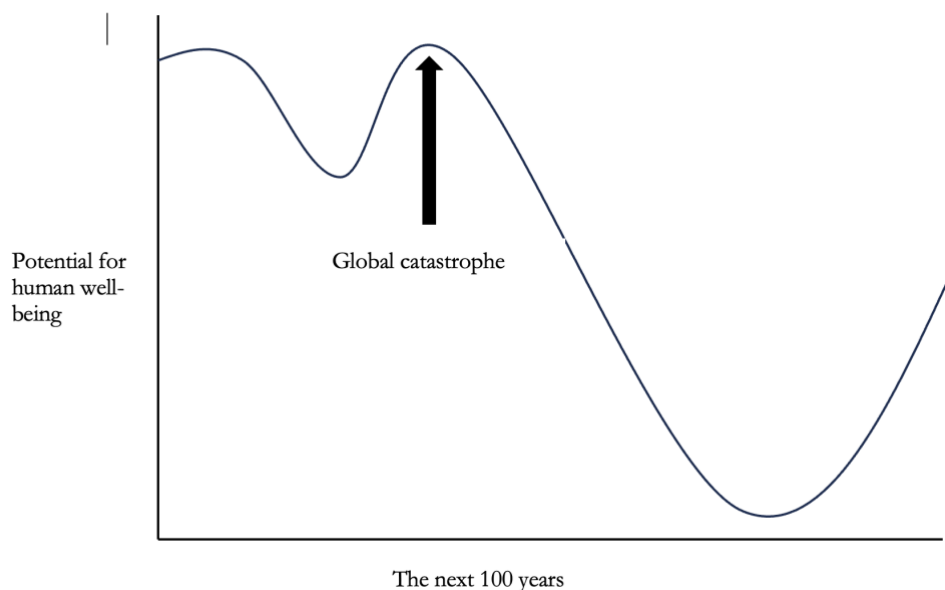
or shifts. I think we are currently living in a point of near peak importance in the development of clean energy, where the global financial system could be recalibrated around renewable energy (potentially even a circular economy). If we focus too much on managing immediate shocks to our current system, we may miss this big opportunity.

Thinking in systems also helps us catch the early signals of collapse. The fall of the Roman Empire may have felt sudden to the inhabitants of Rome, but it resulted from decades of systemic problems (hyperextension of military forces, a burdensome taxation system that created widespread discontent, and an overreliance on slave labour that left the ruling class susceptible to small shocks). Fifty years ago, a report called *Limits to Growth* anticipated that a global collapse would occur in the mid-twenty first century as our energy production became unsustainable.[47] Are we already in the midst of a cascading system failure caused by climate change?[48]

Most of the major proposals we see today aim at building resilience to systemic shocks. This includes the proposals for a Global Resilience Council, or the many ideas about improving preparedness for global disasters.[49] This kind of resilience could help us manage the next crisis, but it can also lock us into dangerous trajectories and obscure deeper risks until it's too late. The carbon-based capitalist system that has driven us to the brink of environmental catastrophe is extremely resilient, surviving even the reality that renewable energy is less expensive today. When we bailed out big banks in 2008, we also contributed to the resilience of that carbon-based financial system. We may have gotten beyond the immediate crisis but potentially baked in a much bigger one.

Could our efforts to build the resilience of our current socioeconomic system lead us into a dystopian future where we have accidentally preserved a world that almost none of us wants? Could the continual band-aids we put on our financial, security, environmental, and other systems be obscuring a slide towards global catastrophic risks that will imperil future generations? We need a better way to evaluate those risks.

(2) A Hive Mind: Diverse Decision-Making

Ironically, those living in democracies tend to think of authoritarian powers as more effective at meeting many global challenges. Western leaders often seem to look wistfully at China, noting how easy it is to make policy when you don't have to face a bothersome electorate. One expert told me: "If China wants to go green, only one person needs to say go."

There is a clear appeal for a centralized body to deal with massive future risks. A world government might cut through the messiness of consensus and distributed decision-making. And such bold action can feel necessary at times of emergency. Indeed, when governments declare states of emergency it usually means a suspension of distributed, democratic decision-making and a temporary consolidation of power.[50]

But the research on this is clear: across a wide range of arenas, a diverse group of amateur problem-solvers will arrive at better solutions than a small group of experts.[51] In global governance settings too, the evidence shows that decisions made by a diverse set of actors are more likely to avoid long-term risks than centralized authoritarian ones.[52] It turns out that there is a strong relationship between citizen participation and effective long-term risk management.[53] More elegantly: "There is an intimate and neglected relationship between existential risk and democracy."[54] A note of caution here: The term "democracy" is a loaded one (in fact the UN avoids it). But at its broadest, the idea of diversity in decision-making would avoid a concentration of decision-making in the hands of a few when it comes to decisions about humanity.

Despite this clear finding in favour of diversity, today a tiny number of actors decide how the world will respond to the most important risks facing humanity. The field of existential and global catastrophic risk studies is dominated by Western academics and overwhelmingly influenced by techno-utopians who believe that we can achieve our maximum potential as humankind via technology.[55] This doesn't mean their views are wrong. In fact, arguments in favor of maximizing human potential for as many people as possible should be persuasive for people across socioeconomic backgrounds.[56] And those who argue that emerging technologies should be aligned with human values make a good case that we should focus on the long-term wellbeing of humanity.[57]

But what are those "human values"? Even amongst philosophers, less than 25 per cent believe in the kind of utilitarianism that guides most of today's longtermism.[58] The vast majority of people have never heard of these arguments, and they certainly aren't making day-to-day decisions with trillions of future humans living on Saturn in mind. It is a safe bet to assume that most people would prefer to avoid global nuclear winter or a pandemic that eradicated humanity. But it is <u>not</u> safe to assume that most people want the kind of future described by elite Western philosophers.

When we look back at our recent history, the majority of the most dangerous decisions for humanity were made by a small group of men. Remember the American scientists willing to light the Earth's atmosphere on fire to beat Hitler to the bomb? Or the Russian submarine captain who defied orders to fire a nuclear weapon on the US during the Cuban missile crisis? Or Putin's casual remarks about the use of nuclear weapons in Ukraine? I can see the irony of myself, a white male, writing about the

need for diversity in decision-making. But it is exactly because I feel so fundamentally unqualified to make these kinds of decisions alone that I am calling for greater diversity in decision-making.[59]

(3) Braking and Decoupling When Risks Grow

Tom Friedman's *Age of Accelerations* rightly points out that most of today's global risks are moving faster than our efforts to contain them.[60] The rapid evolution of AI, climate change, biotech, and even social media are testament to the risks of accelerating change combined with business-as-usual global governance. In many of these areas, we need a way to put the brakes on, preventing exponential changes until better guardrails are put in place. This is one of the key questions facing the newly formed UN AI governance advisory body that will produce a report this year.

Here, complex systems thinking offers two potential braking mechanisms: negative feedback loops and compartmentalization.

Complex adaptive systems operate by feedback loops, responding to new information that is transmitted back into them. Whether ants touching antennas, neurons firing in our brains, or mushrooms passing nutrients, all complex systems adapt to information flows. Negative feedback acts as a brake on runaway change, like the sweat of our bodies prevents runaway overheating.

In managing global catastrophic risks, we need to feed early signals of systemic failure and collective risk back into the system. Some experts call this 'epistemic security,' a way of knowing the world that helps us identify common risks and avoid tipping points.[61] A feedback loop would need to be designed to avoid crossing tipping points and preventing cascade effects. Imagine, for example, a live carbon emissions tracking system that instantaneously sent a public warning when a country or region crossed an emissions red line. Or a social media tracking system that issued an alert when polarized language began crossing a threshold into hate speech and violent rhetoric. This is not completely new thinking. Examples from areas as diverse as crowd control, traffic regulation, and counter-terrorism demonstrate the effectiveness of information feedback and system design in preventing unwanted cascade effects.[62]

Once we reach a tipping point, we then need the ability to prevent the snowball effect. The mycelium mushrooms connecting tree roots detects disease in one tree and stops its spread to others – a form of compartmentalization. Our brains have highly segregated areas for cognitive functions like memory, emotion, and sensory perception. If one area is harmed, the others are often able to carry on by disconnecting from the faulty region. In our tightly interconnected financial system, different mechanisms for 'decoupling' banks help to prevent failing parts of financial systems from infecting others.[63] In the world of cybersecurity, there are ways to quarantine computers when a virus is detected. There are many ways we can think of slowing down and compartmentalizing our global systems to prevent runaway change.

(4) Transformational Governance

If decoupling and braking aren't enough, we move into the realm of transformational governance. Transformation occurs when (a) a system reaches a threshold with unknown or undesirable consequences, and (b) the usual mechanisms of adaptation are not enough to prevent collapse.[64] For example, when some lakes become overly polluted and lose sufficient biodiversity, they shift from a clear body of water to a murky, algae-dominated one. The new equilibrium in the lake involves a completely different relationship amongst the lake's inhabitants and may take some time to settle. This lake has undergone systemic transformation.[65]

Like lakes, social systems can also transform. Our transition to a digitally connected society has occurred without much of a conscious governance process, and we may be on the brink of another

transformation due to artificial general intelligence in the near future. But faced with massive global risks, we should not wait passively like a lake for the transformation to occur. As one leading expert, argued, deliberate transformation requires "radical, systemic shifts in deeply held values and beliefs, patterns of social behavior, and multi-level governance and management regimes."[66]

In global governance, systemic transformation is often about agenda setting and reframing issues, shifting one or more deeply held belief or assumption about the world.[67] This may sound daunting, but it does not necessarily require everything to change.[68] Think of the moment you became aware of the destructive power of a nuclear weapon for the first time, or first really imagined what 8°C of global warming might look like, or thought of a world where AI took most of your decisions for you. For me, these moments caused an almost instantaneous realignment of my worldview, and I continue to think about these scenarios throughout much of my work.

I think the COVID-19 pandemic has pushed us into an era where a critical mass of people woke up to planetary risk. We all felt the contagion cascade as our social, political, and economic systems became overwhelmed.[69,70] This does not necessarily mean the transformation is a good one. In fact, there are worrying indications that our pandemic response has done little to increase our collective willingness to confront these kinds of humanity-level threats. And it may well have baked in some of the worst trends towards global inequality and distrust.

What might a transformational global governance approach look like at the Summit of the Future? We could start with some framing and agenda setting. Imagine if the Pact for the Future adopted at the Summit included reference to the **'rights of future generations.'** This normative reframing of the question of the future could be incredibly powerful, creating a cascade across many other systems. If future generations have rights, all kinds of things change: we need to rethink our approach to GDP as a measure of progress; we may need to have a discount rate to ensure that today's gains are balanced by tomorrow's losses; and we may need to revamp our legal architecture to allow for challenges to environmental harm, or indeed the possession of weapons of mass destruction.[71] The Summit of the Future could be a normative moment that caused much deeper change over time.

Human-Proofing Global Governance

What specific steps could be taken to 'human-proof' global governance? Again, by 'human-proof' I mean a design architecture that combats our innate tendency to prefer the present, undervalue big systemic risks, and ignore the needs of future generations.

There are lots of examples of human-proofing design in today's world. Many hotels require us to put our key into a slot to turn on the lights (combating our laziness when it comes to save electricity). Some governments have an opt-out organ donor box for driving license applications (taking advantage of our laziness for ticking boxes to increase organ donor rates). Restaurants in many parts of the US are required to post bright letter grades in their windows (leveraging our psychology around public shaming to improve restaurant cleanliness). All of these initiatives draw on the idea that we have free will, but that our collective behaviour can be shaped through conscious design.[72]

How could this idea of design be applied to global governance of large-scale future risk? Here are a few ideas, all of which could be part of the lead-up to the Summit of the Future:

1. **Isolating us from presentism.** Governments should invest in spaces, positions, and processes that are not susceptible to the pressures of the present. This could include agreement to non-political appointment of future generations envoys in every government, with dedicated capacities to feed recommendations across different ministries. Or Member States could agree to some long-term plans that could not be altered by election cycles, such as

commitments on investing in sustainable energy. Indeed, the Summit of the Future in 2024 could articulate some of the key areas that should be 'future proofed' by national governments.

2. **A 'gap report' on global catastrophic risk.** One of the most effective tools in recent years has been the United Nations Environment Programme's *Emissions Gap Report*, an important visualization of the gap between national emission commitments and reality. The UN produces a global catastrophic risk 'gap' report offering a similar visualization of the gap between massive future risks and our preparedness. There is a strong psychological impact of seeing a gap between our goals/commitments and where we are today.

3. **Invest in capacities for the future.** Most countries have made public commitments to spending a portion of their gross national production on overseas aid (usually in the range of 0.7 per cent). This public commitment works to keep pressure on governments to maintain a baseline of spending on aid (though many have dipped below the agreed threshold during economic downturns). A similar commitment could be made for future generations at the Summit of the Future, setting aside funds for long-term research, scenario-based planning, and/or agent-based modeling of the future.[73]

4. **The rights of future generations.** One of the ways societies evolve is through the progression of law. In many cases law lags far behind a majority opinion, but over time there tends to be a converging alignment between law and social values. In the 75 years since the signing of the Universal Declaration of Human Rights, prohibitions on torture and enslavement have helped to consolidate societies views on these issues. The Summit of the Future would be the ideal forum to articulate a set of rights for future generations, which could cascade across our governance systems in much the same way as human rights did 75 years ago.

5. **Leapfrog the present.** I often think of Africa's 'leapfrog' of landline technology for phones, allowing many communities to avoid decades of gradual improvements to landline connectivity. To 'future proof' global governance, I believe it will be necessary to do more than gradually inject future-oriented thinking into our work. We need to think bigger, and we probably need to leapfrog many of today's technologies and incremental mindsets. What might a leapfrog idea look like? This would be an ideal topic for a diverse group of young people, but here are a couple of 'stretch' ideas: Could we treat future generations like we treat particle physics, building a multinational scientific endeavour where the world's top scientists jointly develop ideas and technology to safeguard humanity (a CERN for the future, or what Dirk Helbing has called a "planetary immune system")?[74] Or maybe the environmental movement needs to reconsider its aversion to nuclear energy and more seriously consider the emerging science around Thorium-based nuclear power. The point here is to break out of small-scale reforms and think in transformative ways about change.

These are just some starting ideas, drawn from dozens of conversations I've had in recent years. They seek to address a chronic problem in today's world: we are all stressed, busy, rushing from one thing to the next. We don't seem to have time for the kind of open, deliberative reflection needed to transition to a new paradigm. We all struggle to find the "certain slowness" that could enable a deeper reflection on the future.[75] We need to design an architecture for this reflection to drive global governance, or we will be forever chasing the next crisis.

[1] T. Ord, *The Precipice: Existential Risk and the Future of Humanity* (London: Hatchette Book Group, 2020), 167. For a direct critique of this calculation – and indeed Ord's overall approach to existential risk, see, C. Cremer & L. Kemp, "Democratising Risk: In Search of a Methodology to Study Existential Risk," *arXiv* (2021): 11214.

[2] M. Rees, *Our Final Hour: A Scientist's Warning: How Terror, Error, and Environmental Disaster Threaten Humankind's Future in this Century—On Earth and Beyond* (New York: Basic Books, 2003), 8 (putting the odds "no better than fifty-fifty that our present civilization on Earth will survive until the end of the present century").

[3] For some descriptions of these dystopian scenarios, see H. Lin, "The existential threat from cyber-enabled information warfare," *Bulletin of the Atomic Scientists 75*(4) (2019): 187-196; P. Edwards, "Is Climate Change Ungovernable?" Proceedings of the Stanford Existential Risks Conference 2023, 133-146.

[4] HLAB.

[5] J. Boston, "Assessing the options for combatting democratic myopia and safeguarding long-term interests," *Futures* 125 (2021): 102668. But see, A. Karnein, "What's wrong with the presentist bias? On the threat of intergenerational domination," *Critical Review of International Social and Political Philosophy* 26(5) (2023): 725-746 (arguing for a narrow view on future generations focused on avoiding their domination).

[6] J. Boston, *Safeguarding the Future: Governing in an Uncertain World* (Wellington, NZ: Bridget Williams Books, 2017), 8.

[7] This is sometimes called "Knightian Uncertainty," after F. Knight, *Risk, Uncertainty and Profit* (Houghton Mifflin, 1921).

[8] T. Ord, R. Hillerbrand, and A. Sandberg, "Probing the Improbable: Methodological Challengers fro Risks with Low Probabilities and High Stakes," *Journal of Risk Research* 13(2) (2010): 191-205.

[9] T. Ord, *The Precipice: Existential Risk and the Future of Humanity* (London: Hatchette Book Group, 2020), 61.

[10] P. Slovic, "'If I look at the mass I will never act': Psychic Numbing and Genocide," in *Judgment and Decision Making* 2.2 (2007): 79-95; A. Bandura, "Impeding ecological sustainability through selective moral disengagement," *International Journal of Innovation and Sustainable Development* 2(1) (2007): 8–35.

[11] Quoted in J. Lepore "The Atomic Origins of Climate Science," *The New Yorker* 30 (20 January 2017).

[12] N. Bostrom, T. Douglas, and A. Sandberg, "The Unilateralist's Curse and the Case for a Principle of Conformity," *Social Epistemology* 30(4) (2016): 350-71.

[13] For an excellent critique of our flawed distinction between short- and long-term, see M. MacKenzie, "There is no such thing as a short-term issue," *Futures* 125 (2021).

[14] T. Weiss, What Happened to the Idea of World Government, *International Studies Quarterly*, Volume 53(2) (2009): 253–271, 255.

[15] J. Wiener, "The Tragedy of the Uncommons: On the Politics of Apocalypse," *Global Policy* 7 (2016): 67-80.

[16] M. Stauffer, et al., "Hazards with Escalation Potential: *Governing the Drivers of Global and Existential Catastrophes*," United Nations Office for Disaster Risk Reduction (Geneva, Switzerland, 2023) (finding that the UN system is poorly prepared for global catastrophic risk).

[17] See, T. Rowe and S. Beard, "Probabilities, Methodologies and the Evidence Base in Existential Risk," Working Paper, Centre for the Study of Existential Risk, July 2018, available at: http://eprints.lse.ac.uk/89506/.

[18] The most famous iteration of this argument is in J. Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), 287.

[19] For a good overview of this issue, see, W. MacAskill, *What We Owe the Future* (London: Basic books, 2022).

[20] S. Caney, "Political institutions for the future: A five-fold package," in I. González-Ricoy & A. Gosseries (eds.), *Designing institutions for future generations: An introduction* (Oxford: Oxford University Press, 2016), 135–155. See also, A. Jacobs & J. Matthews, "Do citizens discount the future? Public opinion and the timing of policy consequences," *British Journal of Political Science, 42*(4) (2012): 903–935.

[21] For some of the calculations on future populations, see H. Greaves & W. MacAskill, "The case for strong longtermism," Global Priorities Institute Working Paper No. 5-2021, June 2021, 6.

[22] W. MacAskill, *What We Owe the Future* (London: Basic books, 2022); T. Ord, *The Precipice: Existential Risk and the Future of Humanity* (London: Hatchette Book Group, 2020), Appendices A and B; N. Bostrom, "Existential Risk Prevention as Global Priority," *Global Policy* 4 (2013): 15–31.

[23] A Pascalian choice problem is defined by a miniscule probability of something either very good or very bad. For example, a tiny chance of very serious harm to ten trillion people one thousand years from now. See, C. Tarsney, "The epistemic challenge to longtermism," *Synthese* 201 (2023): 195.

[24] See, D. Thorstad, "The scope of longtermism," Global Priorities Institute, Working Paper 6, June 2021.

[25] For an interesting discussion of the value we give to human lives and existential risk, see, I. Raval, "An Axiology of Aesthetics for Existential Risk," Proceedings of the Stanford Existential Risks Conference 2023, 91-103.

[26] I believe the detailed discussions of global public goods are interesting but tend to distract from the most important underlying issues involved in existential risk. I use the term mainly to highlight the collective interest we have in the survival of humanity. Readers interested in the deeper concepts of so-called GPGs should follow Inge Kaul's excellent work at I. Kaul, I. Grunberg, and M Stern, *Global Public Goods* (New York-Oxford, 1999).

27 See, N. Bostrom, "Existential Risk Prevention as Global Priority," *Global Policy* 4(1) (2013): 15-31.

28 K. Ambachtsheer, "The Case for Long-Termism," *Rotman International Journal of Pension Management* 7(2) (2014): 6-15; see also, N. Sears, "International Politics in the Age of Existential Threats," *Journal of Global Security Studies* 6(3) (2021).

29 See, Future of Life Institute (2014), Oxford University's Future of Humanity Institute (2005), the Centre for the Study of Existential Risk at Cambridge University (2012), the Stanford Existential Risks Initiative (2019), and the Käte Hamburger Centre for Apocalypse and Post-Apocalyptic Studies at Heidelberg (2019).

30 D. Zimmer, T. Undheim, P. Edwards, eds., "Intersections, Reinforcements, Cascades," Proceedings of the 2023 Stanford Existential Risks Conference, Stanford Existential Risks Initiative, available at https://doi.org/10.25740/pn116pv4512;

31 A. Einstein, "A Reply to the Soviet Scientists," *Bulletin of the Atomic Scientists* 4(2) (1948): 35-8, 37.

32 A-M. Slaughter, A New World Order (Princeton, N.J.: Princeton University Press, 2004), 8.

33 N. Bostrom, "What Is a Singleton?" *Linguistic and Philosophical Investigations* 5(2) (2006): 48-54.

34 See, e.g., F. Biermann & S. Bauer, *A World Environment Organization: Solution or Threat for Effective International Environmental Governance?* (London: Routledge: 2005); A. McMillan, "Time for a World Court for the Environment," International Bar Association (November 2019) available at https://www.ibanet.org/article/71B817C7-8026-48DE-8744-50D227954E04.

35 D. Wernli, et al., "Understanding and governing global systemic crises in the 21st century: A complexity perspective," *Global Policy* 14.2 (2023): 207-228.

36 C. Rios Rojas, et al., "Building a Science-Policy Interface for tackling the Global Governance of Catastrophic and Existential Risks," University of Cambridge Centre for the Study of Existential Risk, 2023.

37 T. Ord, *The Precipice: Existential Risk and the Future of Humanity* (London: Hatchette Book Group, 2020).

38 D. Bressler & J. Alstott, "The International Panel on Global Catastrophic Risks (IPGCR)," Proceedings of the Stanford Existential Risks Conference 2023, 233-247.

39 Our Common Agenda Policy Brief 2, Strengthening the International Response to Complex Global Shocks – An Emergency Platform, March 2023, available at https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-emergency-platform-en.pdf.

40 Our Common Agenda, Report of the UN Secretary General, 2021, available at https://www.un.org/en/content/common-agenda-report/assets/pdf/Common_Agenda_Report_English.pdf.

41 K. Robinson, *The Ministry for the Future* (New York: Orbit Books, 2020).

42 The landing page for the Welsh Envoy for Future Generations is available at: https://www.futuregenerations.wales/news/wales-leading-the-way-with-future-generations-legislation-un-plans-to-adopt-welsh-approach/.

43 See, J. Boston et al., eds., *Future-Proofing the State: Managing Risks, Responding to Rises and Building Resilience* (Canberra: Australian National University Press, 2014); J. Boston, "Assessing the options for combatting democratic myopia and safeguarding long-term interests," *Futures* 125 (2021): 102668.

44 O. Young, Grand Challenges of Planetary Governance: Global Order in Turbulent Times (Northamptom, MA: Edward Elgar Publishing, 2021), 12.

45 An excellent overview of these characteristics based on water usage examples, is at J. Gupta & S. Schmeier, "Future proofing the principle of no significant harm," *International Environmental Agreements: Politics, Law and Economics* 20 (2020): 731-47.

46 M. MacKenzie, "There is no such thing as a short-term issue," *Futures* 125 (2021), 5.

47 A. Mckay, et al., "Exceeding 1.5 degrees C global warming could trigger multiple climate tipping points," *Science* 377 (2022).

48 S. Galaitsi, et al., "The need to reconcile concepts that characterize systems facing threats," *Risk Analysis* 41 (2021): 3-15.

49 D. Chandler, *Resilience: The Governance of Complexity* (Oxon, UK: Routledge, 2014).

50 J. Ferejohn & P. Pasquino, "The Law of the Exception: A Typology of Emergency Powers," *International Journal of Constitutional Law* 2 (2004): 210–239; L. Kemp, "The 'Stomp Reflex': When Governments Abuse Emergency Powers," BBC Future (2021).

51 L. Hong & S. Page, "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers," Proceedings of the National Academy of Sciences 101 (2004): 16385–16389.

52 H. Landemore, *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many* (Princeton University Press, 2017).

53 J. Sanchez-Ruiz, "Innovative Citizen Participation and New Democratic Institutions: Catching the Deliberative Wave," OECD, 2021, available at: https://www.oecd-ilibrary.org/governance/innovative-citizen-participation-and-new-democratic-institutions_339306da-en.

54 C. Cremer & L. Kemp, "Democratising Risk: In Search of a Methodology to Study Existential Risk," *arXiv* (2021): 11214.

[55] For a good description of the appeal of techno-utopianism, see L. Holt, "Why shouldn't we cut the human-biosphere umbilical cord?" *Futures* 133 (2021): 102821.

[56] For a good summary of the ethical considerations involved in this population argument, see H. Greaves & T. Ord, "Moral Uncertainty about Population Axiology," *Journal of Ethics and Social Philosophy* 12(2) (2017): 135-167.

[57] See, B. Christian, *The Alignment Problem: How Can Machines Learn Human Values?* (New York: Atlantic Books, 2021).

[58] C. Cremer & L. Kemp, "Democratising Risk: In Search of a Methodology to Study Existential Risk," *arXiv* (2021): 11214, 6.

[59] This is not an entirely new idea. See Alvin Toffler's notion of "anticipatory democracy" in A. Toffler, *Future Shock* (New York: Random House, 1975).

[60] T. Friedman, *Thank you for being late: An optimist's guide to thriving in the age of accelerations (Version 2.0, With a New Afterword)* (Picador USA, 2017).

[61] E. Seger, Elizabeth, "Epistemic Security Be a Priority GCR Cause Area?" Proceedings of the Stanford Existential Risks Conference 2023, 18-37. A similar idea of a "knowledge accelerator" is put forward by D. Helbing, Thinking Ahead: Essays on Big Data, Digital Revolution, and Participatory Market Society (Cham, Switzerland: Springer Press, 2015).

[62] D. Helbing, et al., "Saving human lives: What complexity science and information systems can contribute," *Journal of Statistical Physics* 158 (2015): 735-781.

[63] See, M. Guillén, *The Architecture of Collapse: The Global System in the 21st Century* (Oxford Academic, 2015), 51.

[64] B. Chaffin, et al., "Transformative Environmental Governance," *Annual Review of Environment and Resources* 41 (2016): 399–423, 407.

[65] A. Motter & Y. Lai, "Cascade-based attacks on complex networks," *Physical Review E* 66 (2022): 065102.

[66] F. Westley, et al., "Tipping toward sustainability: emerging pathways of transformation," *Ambio* 40 (2011):762–80, 762.

[67] A. Garmestani, "Sustainability science: accounting for nonlinear dynamics in policy and socialecological systems," Clean Technologies and Environmental Policy 16 (2014):731–38.

[68] B. Chaffin, et al., "Transformative Environmental Governance," *Annual Review of Environment and Resources* 41 (2016):399–423, 403.

[69] D. Wernli, et al., "Understanding and governing global systemic crises in the 21st century: A complexity perspective," *Global Policy* 14.2 (2023): 207-228, 211.

[70] L. Böttcher, J. Nagler, and H. Herrmann, "Critical behaviors in contagion dynamics," *Physical Review Letters* 118 (2017): 088301.

[71] Our Common Agenda Policy Brief 4 Valuing What Counts: Framework to Progress Beyond Gross Domestic Product, May 2023, available at: https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-beyond-gross-domestic-product-en.pdf.

[72] See, I. Bohnet, *What Works: Gender Equality by Design* (Harvard University Press, 2016). For a general "future proofing" design framework, see M. Stauffer et al., The FAIR Framework: A Future-Proofing Methodology, Simon Institute for Longterm Governance, April 2023, available at: www.simoninstitute.ch/blog/post/the-fair-framework-a-future-proofing-methodology/.

[73] For some examples, see, E. Lin-Greenberg, R. Pauly, and J. Schneider, "Wargaming for Political Science Research," Social Science Research Network, SSRN Scholarly Paper, February 17, 2021; S. Funnell & P. Rogers, *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models* (San Francisco, Calif.: Wiley, 2011); S. Hyde, "Experiments in International Relations: Lab, Survey, and Field," *Annual Review of Political Science*, 18(1) (2015): 403–424; S. Nowak, et al., *A General Agent-Based Model of Social Learning* (Santa Monica, Calif.: RAND Corporation, 2017)

[74] D. Helbing, Creating ("Making") a Planetary Nervous System as Citizen Web. In: Thinking Ahead - Essays on Big Data, Digital Revolution, and Participatory Market Society (Springer, Cham 2015).

[75] See, F. Cilliers, "What can we learn from a theory of complexity?" *Emergence* 2 (2000): 23–33; K. Rogers, et al., "Fostering complexity thinking in action research for change in social–ecological systems. *Ecology and Society* 18(2) (2013): 31.