Matt Boyd and Nick Wilson

Anticipatory Governance for Preventing and Mitigating Catastrophic and Existential Risks

Abstract

The world faces many large-scale risks. We describe these global catastrophic and existential risks and identify some challenges in governing the prevention and mitigation of such risks. We identify that risk reduction activity in Aotearoa New Zealand has not appropriately addressed these threats. On the basis of the challenges identified, we then deduce the desired features and functions of an entity for effectively governing risk reduction approaches. We argue for an entity that is: anticipatory, central/aggregating, coordinating, apolitical, transparent, adaptive and accountable. We offer structural options for such an entity and outline the merits of several options. **Keywords** anticipatory governance, catastrophic risk, decision theory, existential risk, governance, long-termism, risk assessment

The world faces a suite of extreme risks, which separately or in combination entail catastrophic harm. One objective of good governance should be to reduce the probability of catastrophic harm to as close to zero as possible. Anticipatory governance and long-term risk assessment are essential to this goal. Expected harms may be prevented with timely analysis and action. Unexpected harms can be minimised through good decision-making processes, resilience building and adaptive response.

New Zealand has slowly adopted a forward-looking approach to some individual risks, such as climate change (Climate Change Commission, 2021). There are additional opportunities for identifying small wins, embedding longtermist thinking, giving special attention when long-term interests are at risk, and creating and sustaining an enabling environment for sound long-term governance (Boston, 2021). However, the

Matt Boyd researches health, technology and catastrophic risk, has a PhD in philosophy, and is the owner of Adapt Research Ltd. Nick Wilson is a research professor of public health at the University of Otago, Wellington. The authors declare no competing interests; the research was self-funded by the authors.

Covid-19 pandemic demonstrates that large-scale harms can occur unexpectedly soon, with unforeseen ramifications.

In this article we outline the global catastrophic and existential ('extreme') threats to humanity. We discuss challenges to the governance of this category of risk, before outlining some of the New Zealand government's present risk and resilience mechanisms. We argue that these are insufficient, and then detail the desirable features and functions of an entity tasked with governing extreme risks. We evaluate a set of structural options for establishing an apolitical entity in New Zealand tasked with understanding catastrophic risks and overseeing mitigation measures.

Global extreme risks

The Covid-19 pandemic illustrates many of the problems of large-scale risks. First, the threat of coronaviruses was not appropriately understood by governments, and many pandemic action plans (New Zealand's included) focused narrowly on influenza. Second, New Zealand decision makers had not contemplated the most effective measures ultimately deployed, namely border closure and managed quarantine. Some suggest that Covid-19 was a 'black swan', an event that comes as a surprise, has a major effect, and is inappropriately rationalised after the fact (Taleb, 2007). However, coronavirus pandemics had been identified as a 'time bomb' following the emergence of SARS. Indeed, the Covid-19 pandemic was a paradigmatic 'black elephant', a catastrophe that was extremely likely and widely predicted by experts, but ignored or simply unspoken of (Asayama et al., 2021). Furthermore, knowledge of human cognitive biases explains why we ignore these kinds of risks (Gluckman and Bardsley, 2021; Liu, Lauta and Maas, 2020). Overall, and painfully, not only was the pandemic threat known, but we also knew that we would ignore it. Given this systemic failure, we must look to how we might better anticipate and improve the governance of large-scale risks, because greater threats exist.

The set of global catastrophic risks includes: pandemics, bioweapons, laboratory accidents, artificial intelligence (AI), autonomous weapons, nanotechnology, climate change, geoengineering, ecosystem collapse, nuclear winter, supervolcanic eruption, asteroid/comet strike, global agricultural shortfall, creeping totalitarianism, coronal mass ejection, interstellar events, and other, as yet unknown risks (Bostrom and Cirkovic, 2008; Ord, 2020). These global catastrophic risks could all lead to a loss of 10%, or more, of the human population and/or trillions of dollars of damage through foreseen or unforeseen cascades that bring about states of large-scale harm. The threat is probably rising due to technological advance, increasing global interconnectedness, loss catastrophes could happen unexpectedly soon, including deliberate biological events (Sandberg and Nelson, 2020), unexpected climate feedback loops (Masson-Delmotte et al., 2018), rapid advances in AI (Boyd and Wilson, 2020a), nuclear winter (Robock, 2010; Toon et al., 2019), or previously unknown risks (Ó hÉigeartaigh, 2017).

Accumulating scholarship now describes the psychology of existential risk perception (Schubert, Caviola and Faber, 2019), methodological considerations for estimating or quantifying these risks (Beard, Rowe and Fox, 2020), conceptual

A number of challenges exist that may preclude a full and effective approach to governance of global catastrophic risks and existential risks unless there is specific engineering of institutions.

of diversity, component homogeneity and synchronisation, leading to slow accumulating (Liu, Lauta and Maas, 2018) and/or sudden catastrophic failures (Homer-Dixon et al., 2015).

Existential risks are a subset of global catastrophic risks that could lead to the premature extinction of humanity, or the permanent and drastic destruction of its potential (Ord, 2020). Existential risks are unprecedented and would not allow for meaningful recovery. Mitigation might require international cooperation. Uncertain timing, and/or the sheer scale of the mitigation effort required, might and/or necessitate immediate intergenerational efforts. However, it is rare for governments to explicitly address existential risks. For example, nuclear disarmament is pursued, but nuclear winter is not planned for, and 'unsexy' risks, such as human overpopulation (and irreversible natural resource degradation), do not map well onto traditional disciplinary boundaries or governance (Kuhlemann, 2018). Some existential

frameworks to help manage extreme risk (Torres, 2019), and the world's vulnerability to existential threat (Bostrom, 2019). Risk governance should aim to foresee both near and distant catastrophic events, as well as more nuanced, creeping and fragility-inducing factors that can accumulate. Anticipation would allow prioritising action across the suite of risks in proportion to threat and tractability.

Challenges to the governance of extreme risks

A number of challenges exist that may preclude a full and effective approach to governance of global catastrophic risks and existential risks unless there is specific engineering of institutions. Among them are seven key problems, of anticipation, centralisation/aggregation, coordination, politicisation, transparency, adaptation and accountability. We discuss each in turn.

There are a number of barriers to effectively anticipating large-scale risks. These are founded in short-term thinking, inadequate analytic tools, failure of

Anticipatory Governance for Preventing and Mitigating Catastrophic and Existential Risks

imagination, and uncertainty. Governments face intertemporal policy conflicts, but tend to favour the near term over the long term (Boston, 2021; Gluckman and Bardsley, 2021), yet catastrophic risk governance requires foresight and forecasting. Governance must be alert to a wide range of risks, including the risk of a temporal bias towards the present (Boston, 2017). However, some standard tools (e.g., time discounting) don't allow for future value (and therefore true cost-benefit) to be acknowledged. Furthermore, if risks lead to outcomes that are irreversible or where there are no second chances, then the usual 'as low as reasonably practical' approach to risk might be insufficient. Sufficient

term risks involve 'deep uncertainty', which requires a different set of analytic tools from those typically used in government (Kwakkel, Walker and Haasnoot, 2016). Specialised impartial and quantitative expert risk assessment is needed to overcome neglect of 'uncommons' risks, when learning by experience is not possible (Wiener, 2016).

It can be difficult for organisations to appreciate risks outside their domain, and varying risk methodologies make crosscutting comparison difficult. This means that effective leadership and centralised oversight are needed to ensure aggregation of information and prioritisation of resources across the portfolio of extreme risks, which by their nature have an impact

Short election cycles mean that politicisation can obstruct long-term planning and political decisions risk undermining plans that are underway.

anticipation of global catastrophic risks might require new analytic tools that identify risk at the appropriate granularity (e.g., pandemic rather than influenza pandemic) and key states of harmful affairs (e.g., obscured sunlight, electrical failure), no matter what causal cascades led to them. There is a responsibility to more fully imagine what could go wrong; there needs to be a willingness to search for problems, because one blind spot could spell doom. Ultra-rare but catastrophic risks may be neglected due to psychological unavailability, mass numbing and underdeterrence. Thinking is often obstructed by cognitive barriers, such as difficulties with probabilistic thinking, not caring about people we cannot see and not valuing the future. New Zealand's pandemic preparations had not taken a 'what's the worst that could happen?' approach and attention focused only on influenza, not coronaviruses. Red-teaming approaches (which employ independent experts to critically probe plans for weaknesses) might have anticipated how existing plans could fail. Many catastrophic and longon multiple sectors. This is difficult for any sectoral institution, or set of disconnected institutions, to accomplish. Without a process of centralisation and aggregation, risk analysis may fail to identify instruments and policies that can address multiple risks and drivers in tandem (Kemp and Rhodes, 2020), including strategies that account for complex interactions across risks. It will never be sufficient to task individual government departments with managing extreme risk. This is because of a suite of factors that limit their ability to address large-scale, long-term, cross-cutting risks.

- Departments are busy with day-to-day operational needs; this prioritises the present and obscures slow-onset 'creeping problems' (Boston, 2017).
- The siloed nature of government results in attentional deficits to cross-cutting issues that require central and broad analysis, with sufficient imagination and a forward-looking rather than historical perspective on risks.
- Extreme risks must be understood as a set so that prioritisation and resource

allocation across risks can be done (every department thinks their risks are important).

- Issues of global justice should be considered when preparing for global catastrophic risks, but this is beyond the remit of most departments.
- Government faces both exogenous and endogenous (from within) risks, but most government entities are not in the business of monitoring government for endogenous risks to long-term outcomes (ibid.).

Without large-scale coordination there is a tendency for markets to undersupply large-scale global public goods (Beckstead and Ord, 2014). Only governments or international agencies serve as a mechanism to solve social problems by coordinating various interests across sectors and across departments, and balancing multiple needs, including the needs of present and future generations. The problem of coordination is amplified by the lack of global legal regimes in force that grasp the gravity of extreme risks (Boyd and Wilson, 2020b), and the lack of any coordinated global approach to most extreme risks (Ord, Mercer and Dannreuther, 2021).

Short election cycles mean that politicisation can obstruct long-term planning and political decisions risk undermining plans that are underway. A number of present risk assessment activities take place behind closed doors, and in government agencies that are political (e.g., the Prime Minister's Office) or operate in a political context, where authorisation for relevant risk work may not be forthcoming. Barriers may include concerns around official information requests, or the optics of releasing key information (Kibblewhite and Boshier, 2018). Risk analytic entities such as New Zealand's Department of the Prime Minister and Cabinet may also partially suffer some of these limitations. Additionally, politics can be blind to the long term and to particular moral considerations that transcend politics: for example, the potential immense value of intelligent life on Earth if this is unique in the cosmos. Approaches to rare but extreme risks need to be disconnected from the dayto-day political process.

Additionally, political decisions and processes often lack transparency. As we discuss below, this is not conducive to effective risk reduction. Extreme risks can be complex, unprecedented and difficult to assess and address; therefore, government risk assessment processes should pay special attention to them and this attention should be open to peer review, facilitating appropriate critique of, and attempts to reduce, uncertainty. In New Zealand, legal action against the Climate Change Commission in 2021 to address a claimed error of calculation underscores the importance of transparency.

Rapidly advancing human knowledge and technology (which is both the source of and solution to many global catastrophic risks and existential risks) means that risk governance must be adaptive. Humanity is increasingly capable of having an impact on the geological and ecological world. Entering this era of the Anthropocene challenges traditional human institutions, and existing approaches to risk and mitigation may not be appropriate to safeguard the future. Vulnerabilities enhance risk, and these vulnerabilities include poor risk governance structures. Institutions are good at defending their processes rather than critically assessing them. Risk governance must help institutions examine their own risk processes and improve, despite entrenchment of processes and practices. Additionally, human cognitive biases (such as exponential blindness, or near-term direct causal bias) mean analysts may fail to attend to some risks (Liu, Lauta and Maas, 2020). The increasing threat of extreme risk calls for adaptive design of institutions, and actions which cut across traditional governance silos. This is because the complexity of global catastrophic risks is 'overwhelming the organizational logic of the post-war multilateral order' (Kreienkamp and Pegram, 2020). We note that some rigidity is necessary for staying the course on long-term projects, but this persistence can be supported through a common narrative or vision (van Assche, Verschraegen and Gruezmacher, 2021).

A final challenge to governance of extreme risks is that for many cross-cutting threats there is no individual or organisation that has accountability for oversight of the risk. There must be accountability for understanding and approaching extreme risk and there must be representation of those most likely to suffer harm. Inaction poses a moral hazard, where future anonymous people may be most likely to suffer, yet they are voiceless and powerless in any present deliberation (Kuhlemann, 2018).

If we are to protect humanity from catastrophe, wise decisions must be facilitated through a process that overcomes cognitive biases and aggregates information on disparate risks, and risk and resilience advice must be transparent and independent of politics. Any governance structure for global catastrophic risks must

Defence and Emergency Management, 2019), but it retained a large bias towards natural hazards such as earthquakes and tsunamis (rather than strategic or anthropogenic risks). The strategy does not mention global catastrophic or existential risks and is aligned with the international Sendai Framework (UNDDR, 2015), which suffers from the same blindness. We contest that more attention needs to focus on anthropogenic risks, which probably contain most of the total risk (Ord, 2020). These include risks from non-aligned AI, biological threats and nuclear winter, as well as human impacts on climate and ecology. Since these threats are human generated, we have control over the factors

... any entity tasked with improving New Zealand's resilience to extreme risk must be anticipatory, central/aggregating, coordinating, apolitical, transparent, adaptive and accountable ...

have features to help overcome the seven challenges described above so that it can support the functions needed for effective risk mitigation. We now examine the present state of large-scale risk governance in New Zealand.

Extreme risk governance in New Zealand

A report by the United Nations Office for Disaster Risk Reduction (UNSIDR) in 2017 outlined New Zealand's approach to 'designing, conducting and delivering national disaster risk assessment' (UNISDR, 2017). This report noted New Zealand's traditionally 'siloed' approach to risk assessment (e.g., security agencies, local bodies and scientific agencies acting in parallel). Subsequently, a new national risk assessment process and methodology were deployed which focused on natural threats and operational risks to many government entities. The New Zealand National Disaster Resilience Strategy was published in 2019 (Ministry of Civil

that raise and lower the probability that they eventuate.

The 2021 report Uncertain but Inevitable, written by former New Zealand chief science adviser Peter Gluckman and Anne Bardsley (Gluckman and Bardsley, 2021), notes that governments are responsible for keeping people safe and provides an account of how government thinking on risk and resilience has changed in New Zealand since 2014. The national intelligence and risk coordination team within the Department of the Prime Minister and Cabinet delivered a national risk approach. A multi-year workstream culminated in a national risk register, which allegedly includes 'maximum credible' threats (there are apparently 42 risks across the domains of natural hazards, biological hazards, technological hazards, malicious threats and economic crisis). However, this register is not publicly scrutinisable. We note that the risk profile for 'terrorism' was released in partial

summary form following the Christchurch mosque attacks to satisfy a media official information request. The threat was assessed as 'very high' in the wake of the attacks. However, it is unclear what level was determined prior to this tragedy.

Associated with the Department of the Prime Minister and Cabinet's new approach to assessment of nationally significant risks was a 2018 Treasury discussion paper calling for protection of New Zealand's four capitals: human, social, natural and financial/physical. The report recommended improved institutions for risk mitigation, including legislation, (Ord, 2020). In sum, the present New Zealand approach to extreme risks is at risk of politicisation and lacks transparency, accountability, sufficient foresight and imagination.

National risk registers

New Zealand risk governance presently makes use of a secretive national risk register. The UK, on the other hand, publishes a public, although incomplete, risk register. For example, the UK national risk register mentions nuclear attack, but not nuclear winter. Artificial intelligence is mentioned once in passing. Risks are not

... the present state of extreme risk governance in New Zealand is inadequate in the face of the set of catastrophic and existential risks identified above and the seven challenges to effective governance of extreme risks.

governance and operationalisation (Frieling and Warren, 2018).

The Uncertain but Inevitable report concludes that much more institutional transparency and accountability of risk assessment are needed, with external review of the national risk register. There should be an apolitical focus on highimpact risks that overcomes three particular failures: risk identification, assessment and communication; human factors (especially issues of cognitive biases); and policy/ political dimensions. Any entity overseeing this process should also take a global perspective and focus on the impacts of high-risk events, because the causal factors may be uncertain. The report recommends that the Office of the Auditor-General oversees this. Our foregoing discussion clearly concurs with many of these points. However, the focus on 'inevitable' risks is too narrow, and specific omissions include catastrophic risks posing the greatest threat, namely unaligned artificial intelligence, nuclear war/winter and synthetic biology

listed in order of expected utility loss (per annum or otherwise), so prioritisation (which must necessarily include the additional dimensions of neglectedness, tractability and cost-effectiveness) is difficult.

However, national risk registers are not without criticism (Hagmann and Cavelty, 2012). They are often delimited by national boundaries, and take a problem rather than solution-focused approach. There can be spurious scientific precision, usually reliant on historical data, and a lack of discussion of values, or the structural causal mechanisms behind many anthropogenic risks. Furthermore, uncertainty may be interpreted along lines of vested interest. National risk registers therefore downplay political, normative and ethical questions. Finally, the probabilities factored into national risk registers depend on our actions, and a solution-focused rather than reactive posture could significantly alter the risk matrix (ibid.). Additionally, if we consider the likelihood and impact of some

major catastrophes, or truly existential threats, national risk registers also quite possibly omit almost all the risk, given the fat tail of the distribution of impact. Risk registers are probably important, but in their present form are technically inadequate.

A national risk register should be public in substantial form in democratic countries. There are arguments that some highly sensitive content should be redacted to avoid broadcasting security weaknesses, encouraging perverse investments, or adversely affecting international relations. However, the presumption must be towards open government. The public needs to know that the government acknowledges risk and has plans for addressing (or justification for accepting) risk. Transparency is a commitment device: if a risk is broadcast, it must be addressed (or accepted). The decision to accept risk hinges on risk appetite, and the relevant appetite is the risk appetite of the public and other stakeholders (including future generations). The 2018 Treasury report notes the importance of 'a whole-ofgovernment and whole-of-society response ... a multi-stakeholder coordinated approach to risk management and resilience building'; that 'a strong relationship between the public, private and civil society sectors is pivotal' (Frieling and Warren, 2018, p.38). Openness also facilitates crowdsourcing approaches to risks and solutions (Kankanamge et al., 2018), and superforecasting, a key approach to scenarioising the future (Katsagounos et al., 2021).

Decisions about mitigation (or not) need to balance the values of present people, the rights of future generations, and the wider moral significance of the threat. An open risk register would help facilitate research and engagement on civil society's values with respect to extreme risks. Various methods are appropriate to supplement national risk registers, such as citizen surveys, hui, deliberative democracy and citizen juries (Boyd and Wilson, 2018).

Features and functions of an entity for governing extreme risks

The foregoing suggests that the present state of extreme risk governance in New Zealand is inadequate in the face of the set of catastrophic and existential risks identified above and the seven challenges to effective governance of extreme risks. We now summarise the desired features and functions of an entity tasked with anticipatory governance of extreme risk, before offering a set of possible structural solutions in New Zealand.

Desired features

Given the discussion above, it is clear that the entire risk and resilience process must be governed by an entity possessing certain key features. The entity should be:

- anticipatory;
- central/aggregating;
- coordinating;
- apolitical;
- transparent;
- · adaptive; and
- accountable.

The entity should also be capable of taking a global and intergenerational

perspective, and possess imagination. This entity should be responsible for presenting a coherent and thorough representation of the risks, their probabilities, their impacts, the expected annualised utility loss from each, avenues for prevention/mitigation, and roles and responsibilities, and should help facilitate the required institution building to combat extreme risks. This accounting and planning must be based on research evidence and scientific advice, which must be obtained or developed, if not available. Needless to say, any entity performing these important functions must be well resourced. The next section details these functions of a well-resourced governing entity.

Desired functions

A unique mix of functions and expertise, not found within any existing public sector department, is required for a thorough, aspirational assessment and mitigation

strategy to reduce extreme threats. These functions and expertise include the ability to: identify, articulate and prioritise catastrophic risks; engage with stakeholders; advocate for international cooperation; facilitate wise decision making across government; coordinate across government and across sectors (facilitating institutional reflexivity and an external view); deploy a long-termist perspective using appropriate analytic tools and crossgenerational institution building; cultivate expertise on catastrophic risks and longtermism; and focus on, ideally crosscutting, solutions, including improved risk register methodology (see Table 1).

It might be argued that New Zealand's limited global influence might equate to limited impact in preventing and mitigating global catastrophic risks. However, global catastrophic risks will exhibit an origin and a mechanism of scale-up, and, in the case of existential

Table 1: Desired functions for governing to prevent and mitigate extreme risks

Domain	Specific functions of governing to prevent and mitigate extreme risks	
Identification, communication and prioritisation of extreme risks	 Critically review existing national policy and strategy documents, including red-teaming activities. Problem-finding activity (incentivise identification of risks, including risk inherent in present government structures, rather than minimising of risks). Commission an independent review of extreme risks and analysis to determine which risks justify early commitment and which can wait. Determine the likelihood of a range of catastrophes, their potential impact, and the tractability/cost-effectiveness of mitigation efforts, and rank by annualised expected avoidable disutility impact. Focus on risks neglected by other branches of government. Focus on impacts of risks given uncertainty about precipitating events. 	
Stakeholder engagement	 Reach consensus on 'acceptable risk' among stakeholders (including future generations). Recognise that transparency, crowdsourcing and superforecasting are essential aspects of robust risk reduction. Integrate a te ao Māori perspective on long-termism and risk. Consider education on long-term risk. 	
International engagement	 Advocate for international cooperation on extreme risks. Call out global risk factors that could affect New Zealand and other countries (e.g., reckless Covid-19 policies). Actively cooperate with Australia on large-scale mitigation projects. Contribute to research on and development of methods to help solve collective action problems. 	
Facilitating wise decision making	 Develop improved national risk register methodology that overcomes current weaknesses. Support and facilitate better decision making by developing decision-making and prioritisation tools that overcome human decision-making biases. Develop and deploy decision strategies appropriate in situations of deep uncertainty, rare events and 'creeping normalcy', and to protect future wellbeing. Embed insights from institutions such as CSER, FHI, FLI* and others that study catastrophic and existential risks across government and in key prioritisation decisions. 	
Cross-sector and cross- government coordination (facilitating institutional reflexivity/external view)	 Use a prioritisation framework that crosses sectors and government so that evidence is aggregated and actions with the greatest pay-off are prioritised. Nurture structural changes across all government entities that enable the public sector to take an 'anti-fragile' stance. Avoid an excessively hazard-centric approach and focus on a systems-based and resilience-focused approach. Oversee and consider deeply any major government decisions that are 'irreversible'. 	

Anticipatory Governance for Preventing and Mitigating Catastrophic and Existential Risks

Domain	Specific functions of governing to prevent and mitigate extreme risks		
Long-term focus	 Take a long-termist perspective on risk and employ decision tools appropriate for evaluating long-term strategy. Formalise ways to incorporate the interests of future generations in policy and cultivate a concern for the future. Oversee comprehensive and long-term (e.g., 50 years+) catastrophic risk reporting along with possible solutions in public, unredacted form to encourage innovative solutions. Lay the institutional foundations for projects developing immunity from existential risk, some of which ma span decades. Advocate for, and establish, commitment devices to ensure perpetuation of risk mitigation. 		
Cultivate expertise	 Connect and exchange risk analysis across government. Cultivate research within and outside government to fill identified priority knowledge gaps that will materially affect decisions. Fund secondment of New Zealand experts to international organisations such as CSER, FHI and FLI. Foster ethical leadership that has an understanding of recent advances in moral philosophy. 		
Focus on solutions with oversight of operational activities	 Oversee stress testing of existing response mechanisms to risks well beyond historical examples, including cross-government exercises involving ministerial chief executives and sector leaders. Responsibility and accountability for overseeing mitigation measures. 		

* CSER: Cambridge Centre for the Study of Existential Risk; FHI: Oxford Future of Humanity Institute; FLI: Future of Life Institute

threats, the process will affect every last human. Actions by New Zealand would be wise to focus on risks that may originate locally (such as biological or agricultural threats), on mechanisms for resilience against scale-up once threats emerge, and on surviving threats where New Zealand has a relative advantage in ensuring humanity survives – for example, catastrophic pandemics, biological weapon use, nuclear and volcanic winter (Boyd and Wilson, 2021; King and Jones, 2021). Particular priority areas of activity in New Zealand might include:

- resilience building, in general terms, as well as specific preparations for threats where New Zealand has a relative survival advantage;
- determining which global catastrophic risks might plausibly originate in New Zealand, and associated prevention steps;
- collaborative efforts with Australia, especially where the value of cooperation may be greater than the sum of individual mitigation efforts;
- research into imagining realistic worst case scenarios and problem finding that can be shared with the world;
- contributions to fostering a global workforce of extreme risk expertise; and
- increasing overseas development assistance to the agreed 0.7% of GDP to help neighbouring countries build resilience.

Structural options for governance of extreme risk

Preparing for large-scale risks is one key component of safeguarding the future. So it is illustrative to look at steps other countries have taken. We note that all these international examples fall short of providing capability or capacity to undertake the functions in Table 1. Current initiatives to embed foresight and anticipatory governance in other countries have included establishing a futures commissioner, legislation, think tanks, a government office for science, parliamentary committees, long-term reports, a government council on the future, use of a strong precautionary principle, non-government organisations, and horizon-scanning capability (see Appendix). Many of these initiatives do not have sufficient focus on extreme risks; however, they provide examples of possible institutional structures that may begin to form an ecosystem for extreme risk resilience. Independent researchers have recently published a comprehensive plan that could be implemented in the UK. This 'future proof' approach focuses on addressing biological threats, artificial intelligence, improving government extreme risk management processes, and increasing funding for extreme risk research (Ord, Mercer and Dannreuther, 2021).

In New Zealand there is a need for a substantive first step to act as a catalyst for

change and facilitate the required institutional self-reflexivity and subsequent adaptation. Table 2 lists some contender solutions, and whether they exhibit the features desired of an entity to govern extreme risks.

The ideal approach might be an integrated package of measures. However, first steps in addressing extreme risk must be taken. A well-resourced, independent and capable central entity should design (and redesign as necessary) a catastrophic risk mitigation strategy. The structure must resist procrastination, half-hearted measures and future policy reversal (Boston, 2017). It should nurture capability and development of existing policy, processes and institutions (ibid.). It should have an outward focus towards stakeholders and the global community. Finally, it must aggregate advice from a broad range of experts and stakeholders, and therefore be completely transparent to enable peer review.

Importantly, any mitigation approach must avoid disproportionately preparing for narrowly specified risks (e.g., pandemic influenza versus unspecified pandemics or biothreats), and fighting 'the last war' when the next should be sought. Action must be prioritised by an aggregating mechanism and cost-effectiveness analysis across all risks (while investigating new risks). In some cases, existing risk preparation/ mitigation might advisedly be stopped in favour of shifting resources to higher-

Table 2: Possible New Zealand entities for governing to prevent and mitigate extreme risks

Examples of possible structures	Does the structure have the desired features? (anticipatory, central/aggregating, coordinating, apolitical, transparent, adaptive, accountable)	Advantages	Disadvantages
Risk team within the Department of the Prime Minister and Cabinet	Longer-term anticipatory function would need to be developed; at present lacks transparency.	Currently exists (in a complicated form); close to highest-level decision makers; can bring issues to awareness of the prime minister; deals with issues that cut across all other ministries and agencies.	Potential to be used for political ends; contents of the current risk register are secret; tendency towards securitisation rather than openness.
Independent parliamentary commissioner for extreme risks (as per the parliamentary commissioner for the environment)	Currently does not exist. Could satisfy the required features by design.	Could offer independent advice, with wide powers as an office of Parliament; facilitates a clear sense of ownership/responsibility for advising on the issues; could have a legislative mandate to represent specified future-oriented interests, and requirements for full and transparent regular (time- specified) reporting on activities and advice.	External to the core of government; advice could be ignored, as seems to sometimes be the case with advice from the parliamentary commissioner for the environment.
Parliamentary select committee for extreme risk	Inherently political, so not a stand-alone solution; may lack sufficient anticipatory function given election cycles; insufficiently adaptive given MPs' agendas.	May be relatively sustainable (as part of Parliament's structures – institutional DNA) and can run inquiries.	Traditionally, the New Zealand Parliament has not made sufficient use of expert advice; this would need to be addressed.
Commission (e.g., Climate Change Commission or another Commission for the Future as per the 1980s in New Zealand)	Not sufficiently central; may lack accountability.	Can offer independent and potentially depoliticised advice; facilitates a clear sense of ownership for advising on the issues; could have a legislative mandate to represent specified future-oriented interests, and requirements for full and transparent reporting.	External to the core of government, so advice could be ignored; could struggle to investigate all disparate interests. Aggregation and prioritisation oversight probably has to happen centrally.
Well-resourced team in the Office of the Chief Science Advisor (chief risk and futures advisor)	Not sufficiently central; may lack accountability.	There is a specialised skill set in prioritising and decision analysis under deep uncertainty, and hence a specialised entity (rather than expanded existing capabilities) could provide services to all ministries, as, for example, Treasury does; Office of the Chief Science Advisor started some work in this area (Chief Science Advisor, 2016).	Focus of workstream can shift with new government/new advisor, as seen with the 'Understanding Risk' report series ceasing once Peter Gluckman's term ended and government changed.
Ministry for the Future	Ministries operate in a political context; risk that free and frank advice tempered by ministerial expectations; not sufficiently central; not cross-departmental.	May allow for more critical mass of expertise in one setting than the other arrangements detailed in this Table.	Vertical structure of traditional ministries makes cross-cutting work more difficult; likely to be constrained by the minister or political party in charge.
Mandate for the Office of the Auditor- General to oversee risk assessment	Office already exists. Longer- term anticipatory function would need to be developed; could strike difficulty where entrenched processes may lack adaptivity.	Preferred option of former chief science advisor (Gluckman and Bardsley, 2021). Wide powers as an office of the New Zealand Parliament; apolitical and reports to Parliament not the New Zealand government.	Risk assessment and reporting would be one function among many and might not attract sufficient attention; current focus on auditing may obstruct; multidisciplinary approach needs to be developed.
Develop capability within all existing agencies (e.g., chief futures advisors in all ministries)	Not sufficiently central; aggregating/prioritisation mechanism still required; entrenched processes may limit adaptivity.	Overcomes the issue of expertise (i.e., the problem that any new agency would have in developing expertise across all agencies); could make future-orientation a part of everyday business.	Risk that big novel issues like existential risk or poorly understood technological threats get drowned out by a concern for familiar issues, just on a longer-term scale.

impact areas. These decisions are of critical importance because short-term decisions and unreflective inertia can propagate through time.

We note that 'in-house' agencies, even if mandated by law, can be made impotent by a government that is not supportive, as was the case with the Public Health Advisory Committee under a National government (Skegg, 2021). We note that the advice of 'arm's-length' agencies (e.g., Pharmac) can be ignored, or they can be disestablished, by a new government. The relevant entity needs to be close to the prime minister and Parliament and be well resourced, with a legislative mandate for guaranteed ongoing funding as well as specified reporting requirements. There must be a designated leader who is responsible for ensuring that this brief is fulfilled.

The above requirements and challenges lead us to conclude that a newly established entity, led by someone specifically chosen for their understanding of extreme risks, is most desirable. A parliamentary commissioner for extreme risks working in conjunction with a mandated parliamentary select committee could achieve the aims. Importantly, a commissioner would sit at the heart of Parliament, but would not be bound by election and media cycle pressures. The office could be designed from scratch to satisfy the desired features, without legacy entrenchments, and have a circumscribed focus to attend solely to those factors that have the largest potential impact on the lives of New Zealanders, namely extreme risks. However, we acknowledge that other structures in Table 2 could work if specifically designed to satisfy the seven features we identified above.

We further note that the recent Public Service Act 2020 requires every departmental chief executive to publish a long-term insights briefing independent of ministers every three years (starting in 2022), which should cover medium- and long-term risks. The briefings are to be tabled in Parliament. Unlike other countries, New Zealand lacks a surrounding ecosystem of think tanks, universities and large companies developing long-term views on a range of subjects. For these longterm briefings to be done well, to overcome siloed orthodoxy and cognitive biases, support for the chief executives will be needed. A commissioner, answerable to Parliament (with select committee oversight), could be tasked with supporting risk aspects of these processes. The first round of these reports should be written by experienced multidisciplinary teams, and include the possible impacts of extreme risks, as well as a search for as yet unidentified problems. High-level mitigation strategies should be proposed.

Conclusion

The Covid-19 pandemic suggests that historical decisions have led to widespread lack of preparedness to mitigate global extreme risks. Some decisions today may create path-dependent outcomes in the future, exposing societies to unprecedented risk, possibly destroying large amounts of future value. Mitigating some catastrophic risks might be multi-year, multi-decade or multi-generation projects, which, if not started in time, or if not coordinated internationally, will not be able to address the intended risks in time. Working from the premise that any entity tasked with improving New Zealand's resilience to extreme risk must be anticipatory, central/ aggregating, coordinating, apolitical, transparent, adaptive and accountable, we argue for the establishment of a New Zealand parliamentary commissioner for extreme risks, possibly working in direct synergy with a parliamentary select committee. This project will necessarily be trans-generational, and should include risks where New Zealand is especially well placed to provide some immunity for humanity. The issues and solutions described above will likely generalise to many high-income democracies and there is wide scope for collaborative efforts.

References

Asayama, S., S. Emori, M. Sugiyama, F. Kasuga and C. Watanabe (2021) 'Are we ignoring a black elephant in the Anthropocene? Climate change and global pandemic as the crisis in health and equality', *Sustainability Science*, 16 (2), pp.695–701, doi: 10.1007/s11625-020-00879-7

Beard, S., T. Rowe and J. Fox (2020) 'An analysis and evaluation of methods currently used to quantify the likelihood of existential hazards', *Futures*, 115, 102469, doi: 10.1016/j. futures.2019.102469

- Beckstead, N. and T. Ord (2014) 'Managing existential risk from emerging technologies', in Office of the Government Chief Science Advisor (ed.), Annual Report of the Government Chief Scientific Adviser 2014: innovation: managing risk, not avoiding it: evidence and case studies, London: Government Office for Science
- Boston, J. (2017) Safeguarding the Future: governing in an uncertain world, Wellington: Bridget Williams Books
- Boston, J. (2021) 'Assessing the options for combatting democratic myopia and safeguarding long-term interests', *Futures*, 125, 102668, doi: 10.1016/j.futures.2020.102668
- Bostrom, N. (2019) 'The vulnerable world hypothesis', *Global Policy*, 10 (4), pp.455–76, doi: 10.1111/1758-5899.12718

- Bostrom, N. and M. Cirkovic (eds) (2008) *Global Catastrophic Risks*, Oxford: Oxford University Press
- Boyd, M. and N. Wilson (2018) 'Existential risks: New Zealand needs a method to agree on a value framework and how to quantify future lives at risk', *Policy Quarterly*, 14 (3), pp.58–65
- Boyd, M. and N. Wilson (2020a) 'Catastrophic risk from rapid developments in artificial intelligence what is yet to be addressed and how might New Zealand policymakers respond?', *Policy Quarterly*, 16 (1), pp.53–61
- Boyd, M. and N. Wilson (2020b) 'Existential risks to humanity should concern international policymakers and more could be done in considering them at the international governance level', *Risk Analysis*, 40 (11), pp.2303–12, doi: 10.1111/risa.13566
- Boyd, M. and N. Wilson (2021) 'Optimizing island refuges against global catastrophic and existential biological threats: priorities and preparations', Risk Analysis, doi: 10.1111/risa.13735
- Cabinet Office (2021) *Global Britain in a Competitive Age*, London: Cabinet Office

Chief Science Advisor (2016) *Making Decisions in the Face of Uncertainty: understanding risk*, Wellington, https://www.pmcsa.org.nz/wp-content/ uploads/PMCSA-Risk-Series-Paper-1_final_2.pdf

Climate Change Commission (2021) 2021 Draft Advice for Consultation, Wellington, https://haveyoursay.climatecommission.govt.nz/commsand-engagement/future-climate-action-for-aotearoa/supporting_ documents/CCCADVICETOGOVT31JAN2021pdf.pdf

Farquhar, S., J. Halstead, O. Cotton-Barratt, S. Schubert, H. Belfield and A. Snyder-Beattie (2017) *Existential Risk: diplomacy and governance,* Global Priorities Project

Frieling, M. and K. Warren (2018) Resilience and Future Wellbeing, discussion paper, Wellington: Treasury, https://www.treasury.govt.nz/ publications/dp/dp-18-05

Future Generations Commissioner for Wales (2020) *Future Generations Report 2020*, Welsh Government

Gluckman, P. and A. Bardsley (2021) Uncertain but Inevitable: the expert–policy–political nexus and high-impact risks, Auckland: Koi Tu Centre for Informed Futures, https://informedfutures.org/high-impactrisks/

Government Office for Science (2006) *Infectious Diseases: preparing for the future*, London: Government Office for Science

Government Office for Science (2012) *Reducing Risk of Future Disasters*, London: Government Office for Science

Hagmann, J. and M. Cavelty (2012) 'National risk registers: security scientism and the propagation of permanent insecurity', *Security Dialogue*, 43 (1), pp.79–96, doi: 10.1177/0967010611430436

Homer-Dixon, T., B. Walker, R. Biggs, A-S. Crépin, C. Folke, E. Lambin, G.D. Peterson, J. Rockström, M. Scheffer, M. Steffen and M. Troell (2015) 'Synchronous failure: the emerging causal architecture of global crisis', *Ecology and Society*, 20 (3), doi:10.5751/ES-07681-200306

Kankanamge, N., T. Yigitcanlar, A. Goonetilleke and M. Kamruzzaman (2018) 'Can volunteer crowdsourcing reduce disaster risk? A systematic review of the literature', *International Journal of Disaster Risk Reduction*, 35, 101097, doi: 10.1016/j.ijdrr.2019.101097

Katsagounos, I., D. Thomakos, K. Litsiou and K. Nikolopoulos (2021) 'Superforecasting reality check: evidence from a small pool of experts and expedited identification', *European Journal of Operational Research*, 289 (1), pp.107–17, doi: 10.1016/j.ejor.2020.06.042

Kemp, L. and C. Rhodes (2020) *The Cartography of Global Catastrophic Governance*, Cambridge, https://globalchallenges.org/the-cartography-of-global-catastrophic-governance/

Kibblewhite, A. and P. Boshier (2018) 'Free and frank advice and the Official Information Act: balancing competing principles of good government', *Policy Quarterly*, 14 (2), pp.3–9

King, N. and A. Jones (2021) 'An analysis of the potential for the formation of "nodes of persisting complexity", Sustainability, 13, 8161, doi: 10.3390/su13158161

Kreienkamp, J. and T. Pegram (2020) 'Governing complexity: design principles for the governance of complex global catastrophic risks', *International Studies Review*, doi: 10.1093/isr/viaa074

Kuhlemann, K. (2018) 'Complexity, creeping normalcy and conceit: sexy and unsexy catastrophic risks', *Foresight*, 21 (1), pp.35–52, doi: 10.1108/FS-05-2018-0047

Kwakkel, J., W. Walker and M. Haasnoot (2016) 'Coping with the wickedness of public policy problems: approaches for decision making

under deep uncertainty', *Journal of Water Resources Planning and Management*, 142 (3), 01816001

Liu, H., K. Lauta and M. Maas (2018) 'Governing boring apocalypses: a new typology of existential vulnerabilities and exposures for existential risk research', *Futures*, 102, pp.6–19, doi: 10.1016/j. futures.2018.04.009

Liu, H-Y., K. Lauta and M. Maas (2020) 'Apocalypse now? Initial lessons from the Covid-19 pandemic for the governance of existential and global catastrophic risks', *Journal of International Humanitarian Legal Studies*, 11 (2), pp.295–310

Masson-Delmotte, V., P. Zhai, H. Pörtner, D. Roberts, J. Skea, P. Shukla et al. (2018) Global Warming of 1.5°C: an IPCC special report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty, Geneva: IPCC, https://www.ipcc.ch/sr15/

Ministry of Civil Defence and Emergency Management (2019) National Disaster Resilience Strategy, Wellington, https://www.civildefence.govt. nz/cdem-sector/plans-and-strategies/national-disaster-resiliencestrategy/

Mucci, A. (2015) 'Sweden's Minister of the Future explains how to make politicans think long-term', Vice, 26 November, https://www.vice.com/ en/article/ezp4am/swedens-minister-of-the-future-explains-how-tomake-politicians-think-long-term

Ó hÉigeartaigh, S. (2017) 'Technological wild cards: existential risk and a changing humanity', in *The Next Step: exponential life*, Madrid: BBVA

Ord, T. (2020) *The Precipice: existential risk and the future of humanity,* London: Bloomsbury

Ord, T., A. Mercer and S. Dannreuther (2021) *Future Proof: the opportunity to transform the UK's resilience to extreme risks*, London: Centre for Long-term Resilience, https://www.longtermresilience.org/futureproof

Prime Minister's Office (2018) *Government Report on the Future*, Helsinki: Prime Minister's Office

Robock, A. (2010) 'Nuclear winter', *Wiley Interdisciplinary Reviews: Climate Change*, 1 (3), pp.418–27, doi:10.1002/wcc.45

Sandberg, A. and C. Nelson (2020) 'Who should we fear more: biohackers, disgruntled postdocs, or bad governments? A simple risk chain model of biorisk', *Health Security*, 18 (3), pp.155–63, doi:10.1089/hs.2019.0115

Schubert, S., L. Caviola and N. Faber (2019) 'The psychology of existential risk: moral judgments about human extinction', *Scientific Reports*, 9 (1), 15100, doi: 10.1038/s41598-019-50145-9

Scott, D.P. (2016) 'Application of the precautionary principle during consenting processes in New Zealand: addressing past errors, obtaining a normative fix and developing structured and operationalised approach', LLM thesis, Victoria University of Wellingotn

Skegg, D. (2021) 'The Covid-19 pandemic: lessons for our future', *Policy Quarterly*, 17 (1), pp.3–10

Taleb, N. (2007) The Black Swan, Random House

Toon, O., C. Bardeen, A. Robock, L. Xia, H. Kristensen, M. McKinzie, J. Peterson, C.S. Harrison, N.S. Lovenduski and R. Turco (2019) 'Rapidly expanding nuclear arsenals in Pakistan and India portend regional and global catastrophe', *Science Advances*, 5 (10), doi: 10.1126/sciadv. aay5478

Anticipatory Governance for Preventing and Mitigating Catastrophic and Existential Risks

- Torres, P. (2019) 'Facing disaster: the great challenges framework', *Foresight*, 21 (1), pp.4–34, doi: 10.1108/FS-04-2018-0040
- UNDDR (2015) Sendai Framework for Disaster Risk Reduction 2015– 2030, Geneva: United Nations Office for Disaster Risk Reduction, https://www.undrr.org/publication/sendai-framework-disaster-riskreduction-2015-2030
- UNISDR (2017) National Disaster Risk Assessment Words into Action Guidelines Governance System, Methodologies, and Use of Results, Geneva: United Nations Office for Disaster Risk Reduction, https:// www.unisdr.org/files/52828_nationaldisasterriskassessmentwiagu.pdf
- van Assche, K., G. Verschraegen and M. Gruezmacher (2021) 'Strategy for the long term: pressures, counter-pressures and mechanisms in governance', *Futures*, 131, doi: 10.1016/j.futures.2021.102758
- Welsh Government (2019) Wales and the Sustainable Development Goals, Welsh Government, 2019
- Wiener, J.B. (2016) 'The tragedy of the uncommons: on the politics of apocalypse', *Global Policy*, 7 (S1), pp.67–80, doi: 10.1111/1758-5899.12319

Appendix: Selected strategies for embedding foresight and anticipatory governance into government in other countries, illustrating a range of possible approaches

Jurisdiction	Strategy	
Wales (UK)	egislation and commissioner for the future Wellbeing of Future Generations Act 2015 Future generations commissioner with statutory powers to represent people who haven't yet been born Executive summary of the <i>Future Generations Report 2020</i> (Future Generations Commissioner for Wales, 2020) focuses on environment/climate/jobs/transport/wellbeing Report recommends a minister for prevention, and budget for prevention activities across government (see p.22 in the executive summary) Mentions 'catastrophic' with regard to sixth mass extinction/climate, but not other catastrophic risks and doesn't mention 'existential' Only three PDFs mentioning 'catastrophic' on the futuregenerations.wales website; none mention 'existential' 2019 <i>Wales and the Sustainable Development Goals report</i> (Welsh Government, 2019) mentions 'improving resilience to disaster'; specifically mentions flooding and coastal erosion along with the words 'catastrophic risks'	
Scotland (UK)	 Independent think tank Scottish Futures Forum – Scotland Parliament think tank to promote research and stimulate debate (since 2005) Provides content about sustainability, education, future horizons of ten years, e.g. Scotland 2030 programme Focus seems to be narrow and 'short-term': climate change and the future of work to 2030 'and beyond' 	
England/UK	 Government Office for Science Proposed Wellbeing of Future Generations Bill Ministerial oversight from minister for cabinet office Foresight team works on projects where there is a department that will champion them, reports include <i>Reducing Risk of Future Disasters</i> (Government Office for Science, 2012), natural hazards only; and <i>Infectious Diseases: preparing for the future</i> (Government Office for Science, 2006) Horizon-scanning team in the Government Office for Science in London produces research on, e.g., artificial intelligence, demographic change, emerging technologies; fosters communities of interest across civil service Government Office for Science report <i>Innovation: managing risk not avoiding it</i> (Government Office for Science, 2014) includes a chapter on 'managing existential risk from emerging technologies' <i>Global Britain in a Competitive Age</i> integrated review, including ten-year strategy and mention of low-probability but catastrophic risks (Cabinet Office, 2021) National security risk register (but immediate focus: focuses on 'events' not risks, on 'attacks', 'accidents', but not systemic risks; also no peer review. Evidence given to House of Lords Committee on Risk Assessment and Risk Planning, 13 January 2021: The National Risk Register omits the very many ways in which these technologies interact across the board (e.g., steam engine technology leading to railways, modern warfare, and the rise of communism and fascism). The risk register omits many really important risks. By putting technological risks alongside events like 'flooding' it misses an account of undesirable outcomes (e.g., breakdown of transport system, civil unrest, erosion of democracy, etc.) and how these become more likely with e.g., artificial intelligence. 	

Jurisdiction	Strategy	
	 House of Lords call for evidence on national risk assessment and risk planning (expert evidence panels January/February 2021) Well-being for Future Generations Bill (private member's bill, introduced 2020, at second reading stage in mid-2021) 	
Finland	 Parliamentary committee Long-term reports Committee for the Future in Parliament consists of 17 members of the Finnish Parliament, serves as a think tank for futures, science and technology policy in Finland Government 'long-term reports' once each election cycle, 10–20 year focus – e.g., the <i>Government Report on the Future</i> (Prime Minister's Office, 2018) focused on the future of work. The report process always involves Parliament, and aims to encourage broad debate Finnish Ministry for Foreign Affairs supported an Oxford FHI report, <i>Existential Risk: diplomacy and governance</i> (Farquhar et al., 2017) 	
Sweden	 Government Council on the Future Minister for strategic development led 'Mission: the Future' (2014) Council on the Future created by the minister, comprising seven MPs plus the prime minister. There are three working groups: future of work, fossil-free society and global coordination Horizontal coordination across many ministries is important, says the minister: we are an 'internal government think tank whose role is to constantly remind others to include the long-term in the decision making process' (Mucci, 2015) '• Total defence concept' national resilience exercises across 15 national agencies 	
Canada	 Strong precautionary principle Precautionary principle use is stronger than in New Zealand A Supreme Court ruling based on application of the precautionary principle in law is likely to give governments a broamandate to reject anything that 'has potential' to harm the environment Note that New Zealand has exhibited confusion and misapplication when the precautionary principle has appeared in la (Scott, 2016) 	
Australia	Non-governmental organisation Australian Human Futures Commission (fledgling, circa 2020) 	
Singapore	 Advanced and dedicated risk/horizon-scanning capability Risk Assessment and Horizon Scanning Experimentation Centre Uses cognitive (artificial intelligence) tools to aid analysts in identifying threats; cross-government, joining silos Centre for Strategic Futures (2010) – internal think tank 	

High quality professional development courses for the public sector:

CROWN MAORI RELATIONS: TREATY AND MORE

-> Mon 8 & Thu 11 November, 9am-1pm

DELIVERING DATA-DRIVEN INSIGHTS

-> Mon 15 & Tue 16 November, 9am-330pm

ENGAGING EFFECTIVELY WITH YOUR STAKEHOLDERS

-> Thu 18 November 9am-430pm



PUBLIC POLICY FUNDAMENTALS -> Wed 24 November, 9am-430pm

INFORMATION ARCHITECTURE FUNDAMENTALS -> Wed 1 & Thu 2 December, 9am-430pm

ASSESSING CYBER SECURITY RISK

-> Thu 2 & Fri 3 December, 9am–5pm

We can deliver courses in-house, customise or design new programmes to suit your requirements at our Wellington and Auckland teaching spaces.

For information and to enrol visit: www.wellingtonuni-professional.ac.nz or call us on 04-463 6556.