Matt Boyd and Nick Wilson[1]

# Existential Risks

## New Zealand needs a method to agree on a value framework and how to quantify future lives at risk

**Abstract**

Human civilisation faces a range of existential risks, including nuclear war, runaway climate change and superintelligent artificial intelligence run amok. As we show here with calculations for the New Zealand setting, large numbers of currently living and, especially, future people are potentially threatened by existential risks. A just process for resource allocation demands that we consider future generations but also account for solidarity with the present. Here we consider the various ethical and policy issues involved and make a case for further engagement with the New Zealand public to determine societal values towards future lives and their protection.

**Keywords** Existential risk, future lives, public engagement, risk mitigation, value framework

**Matt Boyd** is a philosopher and health researcher and the Director of Adapt Research Ltd, Wellington. Corresponding author: 14 Broadway, Reefton; email: matt@adaptresearchwriting.com; phone: +64223512350. **Nick Wilson** is a Professor of Public Health in the Department of Public Health, University of Otago, Wellington.

Most policy work focuses on present concerns to existing people. Political leaders and public policy workers typically consider benefits over a limited time horizon – such as just the time before the next election. But some social projects involve benefits many decades into the future: public infrastructure (roads, bridges, hospitals, civic buildings), establishing national parks and marine reserves, and establishing treaties such as the Montreal Protocol (on ozone depletion) or the Paris Agreement (climate change).

Sometimes we exhibit concern for the welfare of people beyond our lifetimes. For example, we consider how to store nuclear waste safely over thousands of years. People sometimes consider injustices done to past generations as well, through present-day settlement of claims relating to past treaties, such as the 1840 Treaty of Waitangi in New Zealand.

*Existential risks*

Larger-scale existential risks are events or processes which could cause the extinction of the human species, or end organised human civilisation. These include widespread nuclear war, runaway climate change, biodiversity loss, ecological crises, synthetic bioweapons, superintelligent artificial intelligence run amok, asteroid impacts, and interstellar events such as gamma ray bursts (Bostrom and Cirkovic, 2008; Rockstrom et al., 2009).

There is growing literature on the potential value of preventing existential risks (Bostrom, 2013; Matheny, 2007; Tonn and Stiefel, 2014), along with issues of intergenerational justice (Adler, 2009; Arrhenius, 2000; Arrhenius and Rabinowicz, 2010; Broome, 2005; Gosseries and Meyer, 2009; Meyer, 2018; Narveson, 1967; Tarsney, 2017; Weitzman, 1998). In 2014 the World Economic Forum global risks report made no mention of many human existential risks, yet the 2017 report specifically addresses failures in cooperation on climate change and the threat of weaponised artificial intelligence, and the fact that governing institutions remain reactive and slow moving (World Economic Forum, 2014, 2017).

New Zealand publications discuss some issues of long-term or existential risk management (Boston, 2017; Boyd and Wilson, 2017; Council of the New Zealand Ecological Society, 1985), but there is as yet no coordinated response to existential threats. This is despite the *Bulletin of the Atomic Scientists* announcing that the symbolic Doomsday Clock, representing the threat of human destruction, has recently advanced to two minutes before midnight (Bulletin of the Atomic Scientists, 2018).

In this article we outline several philosophical approaches one might take when valuing future people. We then argue that when there are several coherent positions available to policymakers, we ought to have public engagement and community debate to ensure sustainable policy responses and long-term investments consistent with public views. We explain how this might be done using emerging empirical philosophical strategies. The reason for all this is that if we do value future people, and we are capable of mitigating existential risks, then perhaps we ought to do that.

We then present our own utility calculations for the number of future New Zealand life-years at risk – including when discounting is used – although we note that utility calculations may not be the only important considerations, pending the outcome of the public engagement we describe.

We may value the lives of future generations, and perhaps have obligations towards their well-being, or we can deny that their lives have value.

**New Zealand needs an agreed framework for how we value future lives**

*Valuing future lives*

An important question shaping how we act today is, 'what do we owe to future people?' The answer can range from 'everything' (even to the point of overdemandingness on our own lives and resources) through to 'nothing'. We may value the lives of future generations, and perhaps have obligations towards their well-being, or we can deny that their lives have value. We now describe some different ways in which a society might choose to value human life in the future. We emphasise that it is unclear which view New Zealanders take on average as a population and how diverse these views are.

Some 'person-affecting' views of morality posit that acts can only be wrong if they affect someone, but future people don't presently exist (Parfit, 1984). That said, we have little difficulty grasping the wrongness of imposing risks on future people. A 'risky policy' which results in predictable deaths in 300 years still seems bad, irrespective of who is actually killed (ibid.). It is the predictability of the deaths that is important rather than the actual people who might be killed. These future people would still regret our present decisions. However, it could be the case that we should give lesser weighting to the value of future lives through some rate of discounting or temporal partiality.

*Temporal partiality*

If we think about the present, we may find that we treat different humans differently, here and now, for supposedly legitimate reasons. For example, a person may be praised for spending $100,000 on an operation to save her sister, even though she could have spent $100 to save each of 1,000 starving children. It can be argued that obligations to people diminish with distance and degree of personal relatedness. Close human relationships matter in all societies and this person may not be condemned for saving her sister in this way, even though there was a moral opportunity cost. Heyd articulates a similar idea in terms of 'solidarity' (Heyd, 2008). Such considerations of partiality arise frequently in policy discussions: for example, around aiding refugees versus investing in local people.

It may also be the case that our obligations to distant people diminish similarly with time. This adds weight to the case for some level of discounting of the value of future lives. We may not be condemned if we fail to prevent an extinction event far in the future.

Some strict utilitarians might challenge the woman who committed $100,000 to save her sister, because relatedness and distance should not matter: all human lives should be considered equally valuable and if we can save 1,000 rather than one, we should (Singer, 1972). Such a utilitarian might claim that resources should be used for those in the world in greatest need, right up to the point of marginal utility to the individual with resources available. This is a very demanding conception of morality (Sonderholm, 2013); however, most developed societies demonstrate some level of obligation to distant people through

various assistance programmes. But it is not obvious to what degree we should value people distant in time.

A further problem for the temporal partiality position is that if we are partial towards the present, then it looks like the value of righting past wrongs must also diminish. If temporal partiality in favour of the present is permitted, then, given a symmetrical relationship between past and future, we might be justified in discounting reparations for wrongs of past generations such as slavery, conquest or breach of treaties.

*Actual versus statistical lives*

Sometimes we discriminate in favour of known individuals in present danger rather than statistical lives at risk (Weale, 1979). For example, intensive care units expend heroic amounts of resources on individuals. This is inconsistent with claims that it is generally wrong for a funding organisation to fund individual 'rescue' over mass prevention (Hope, 2001).

Current prevention activities, such as providing clean water and sewerage systems, immunising a population to achieve herd immunity, and taxing alcohol and tobacco, are interventions on a known population, with known statistical pay-offs. Robust research has established the risks and probabilities. However, for existential risks the issue of prevention is more complex, as it may involve intervening with respect to a less well-defined population (future people) for a possible pay-off (the existential threat may or may not occur).

Furthermore, we are more uncertain of the needs of future people. They may be very much more wealthy than we are now, with technology we can't imagine. This uncertainty around the commitment of resources to avoid an existential risk may also justify some discounting of the value of future lives.

However, human life is a qualitatively different kind of good from other resources. This is in part because human lives are not obviously tied to estimates of inflation/depreciation and future value as material goods are. Therefore, there seem to be no good reasons to prefer one discount rate over another. Indeed, most authors writing on intergenerational justice seem opposed to discounting future lives (Matheny, 2007; Gooseries and Meyer, 2009). The consideration of whether to apply a discount rate, and what the rate should be, is important in this context, because the choice among discount rates will have significant implications on calculated value when we are looking far to the future.

*Fairness about existence*

Equity or fairness considerations are often used in conjunction with utility when determining policy. Rawlsian considerations of justice apply a fairness principle and offer us a social contract under a 'veil of ignorance' to illustrate the uncertainty, prior to our existence, of our circumstances (male or female, privileged or not privileged, and so on). According to this argument, we should construct society so that circumstances are fair regardless of who we are (Rawls, 1971). Of course, ignorance applies to *when* we exist as well.

Under such terms, creating a safe environment for everyone presently and maintaining this level of welfare for future people would constitute fair policy. Such considerations have been used to argue for moderation of present resource use and environmental protection (Norton, 1989). If fairness demands that we protect a present person's future life-years irrespective of social circumstances (for example, through healthcare provision), then this ought to apply to future people as well. For example, future people might have a right to a life of natural length. Furthermore, according to some moral frameworks, if it is within our power we might be obliged to ensure future people enjoy levels of well-being at least equivalent to those enjoyed by present people.

In general, to the extent that ethics is impartial, and thus the well-being of one person does not automatically trump the well-being of someone else, then distance in relatedness, location, and perhaps time will lose relevance. Additionally, most ethicists seem to agree that impartiality must be at least some part of ethical thinking. This is because a totally partial ethic is moral egoism (concern only for oneself), and this is not what most people mean by an ethical view. Therefore, we must to some degree consider the well-being of those other than ourselves. It seems prima facie reasonable to posit that this consideration for others might need to extend beyond our own society in order to have fully informed ethical deliberation. We must at least consider lives distant from ourselves when considering the rightness of our own actions, and this perhaps ought to extend to future lives or societies as well.

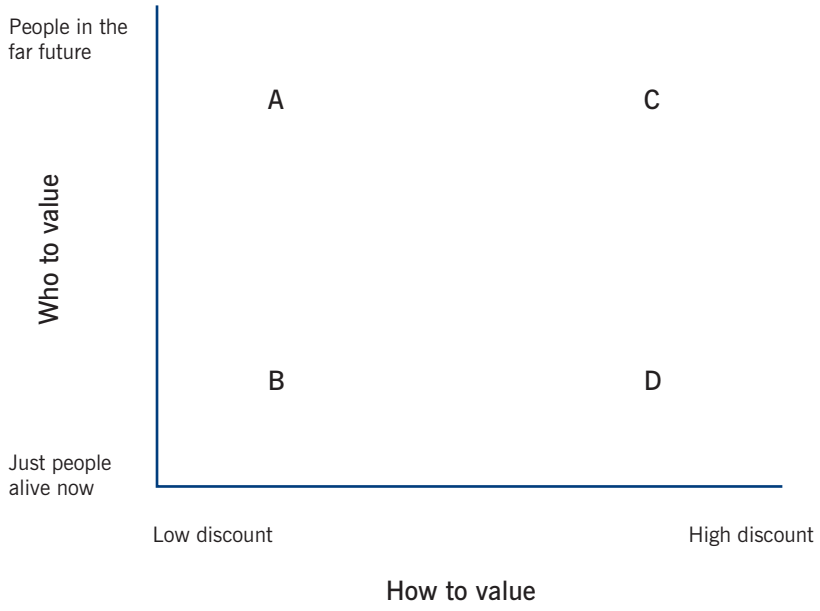
In general, to the extent that ethics is impartial, and thus the well-being of one person does not automatically trump the well-being of someone else, then distance in relatedness, location, and perhaps time will lose relevance.

*Public reason*

Bostrom assumes that it does not matter when a life exists, and therefore we ought to spend vast resources protecting the many billions of future lives (Bostrom, 2013). However, it is exactly these kinds of assumptions that we need to test at the level of the New Zealand population through public engagement. We suggest that the leap from 'future lives matter' to 'future lives matter equally with present lives' needs close consideration. Indeed, Bostrom in fact agrees with this point when he argues:

> In a similar vein, an ethical view emphasising that public policy should be determined through informed

**Figure 1: A concept space for valuing future lives**




Legend: A – future life-years are equally valuable with present; valuing the 'human project' fits in here;
B – only value those presently alive, but all their future years are equally valuable to the present day;
C – future life-years are valuable, but not equally so with those presently alive;
D – only value those presently alive, and favour life-years in the immediate future.

> democratic deliberation by all stakeholders would favour existential-risk mitigation if we suppose, as is plausible, that a majority of the world's population would come to favour such policies upon reasonable deliberation (even if hypothetical future people are not included as stakeholders). (ibid., p.23)

It is exactly the conclusions of 'reasonable [public] deliberation' that we need. Consideration of these issues must precede, and will shape the use of, any discount rate on the value of future lives. Public engagement will help inform policymakers as to which risk mitigation rule is appropriate, especially considering that substantial diplomatic effort and financial resources might be needed to address certain existential risks.

*The human project*

Finally, there is an important distinction between considering future 'people' or future 'life-years' and considering future 'generations'. The latter are critical components of the 'human project', such as the continuity of cultural, scientific and technological endeavours across generations. Humans particularly value these projects (Scheffler, 2013), and the long-term persistence of such projects depends on subsequent generations actually existing.

In particular, Scheffler argues that we need future humans in order that many things can matter to us now. In his view the imminent end of our species would produce widespread 'apathy, anomie, and despair ... and ... a pervasive loss of conviction about the value or point of many activities' (ibid., p.40). If this is accurate, then the existence of people after we die is an important condition of things mattering to us now. While not denying the general importance of self-interested motivation, Scheffler concludes that: 'there is a very specific sense in which our own survival is less important to us than the survival of the human race' (p.73).

We add that, importantly, when considering actual threats of human extinction, by protecting known lives in present danger we are also protecting future lives in potential danger.

*A concept space of value*

In summary, there is a range of positions New Zealand society could take with respect to future lives. These are illustrated in Figure 1. Once we have evaluated the worth of the 'human project', our uncertainty about the future, who is deserving of consideration, and possible discounting of future lives, we will know whether our position as a society is nearer to A, B, C or D.

**We can establish which of these frameworks to apply through public engagement**

As argued above, there is no doubt that humanity faces a range of existential threats. However, it is unclear what action against these threats we should take, given that we can approach the future of humanity from these different philosophical perspectives. Various perspectives may be defensible, and which approach best coheres with the intuitions of New Zealand people is unknown, yet such information should be a critical input into decision making in a democratic society with limited resources (Bromell, 2012; Gluckman, 2011). The process by which decisions about the investment of public resources are made must be a just process (Daniels and Sabin, 1997), and public policy requires us to engage with diverse others in public reasoning (Freiberg and Carson, 2010; Nussbaum, 2000).

In undertaking such deliberation, then, policymaking requires both evidence and morality. Policymakers informed with the best evidence cannot unilaterally decree morality. There is no avoiding the 'normative jungle' in policymaking (Gruen, Kelly and Gorecki, 2011). We need a public exchange of reasons informed by relevant evidence (Rawls, 1987). The research question, 'Which value framework encompassing future people and protection of the human project best coheres with the views of New Zealanders?' needs to be explored.

Recent innovations in philosophy import empirical methods from the social sciences, which many ethicists see as an important adjunct to philosophical enquiry (Kahane, 2013; Tanyi and Bruder, 2014). These methods access ordinary people's intuitions to supplement the investigations of ethicists and philosophers. This synergistic method can bolster philosophical reasoning and offer novel insights, such as previously unrecognised distinctions (Deery, Davis and Carey, 2014).

Some experimental philosophers have gone further than seeking intuitions about

abstract or concrete situations and employ questionnaire scales and statistical techniques such as factor analysis to reveal the structure of survey data collected (ibid.; Nadelhoffer et al., 2014). This methodology has not yet been explored in the domain of future lives and intergenerational justice.

We suggest that New Zealand policymakers are obliged to gather reasoned public opinion, perhaps through the use of key informant interviews, citizen juries, hui, surveys or the like. The aim of public engagement is to access New Zealanders' values and reasoning. Questions, vignettes or discussion topics should aim to access not just judgements about value, but also preferences, given the potentially large opportunity cost of acting to mitigate certain existential risks. Qualitative and quantitative methods could be used to seek reasons behind the intuitions about the value of future lives and the 'human project'.

Utility is currently a central concern to the New Zealand government when setting policy. We see this in areas of health resource allocation or transport safety, for example. Treasury's cross-government CBAx modelling tool prompts for explicit input of utility (New Zealand Treasury, 2017). Although not the only measure of value, utility is likely to remain central to policy decision making. The outcome of public engagement in the domain of existential risk will determine which utility calculations policymakers must undertake when calculating the costs and benefits of investing in prevention of existential threats. (For example, should we calculate the number of life-years at risk of those presently alive, or of all future New Zealanders? And should we apply a discount rate to future life-years or not?) These calculations will determine what level of investment in preventing existential threats is justified. Furthermore, should we decide to invest in mitigation, we know that policies that require some sacrifice are more likely to be adopted successfully following extensive engagement and dialogue with interested and affected parties.

In sum, we need to know which of the philosophical positions outlined (A to D in Figure 1, or variations of them) the New Zealand public actually hold or would support on further reflection; crucially this must include determination of Māori views. We can then supplement the value position with evidence on the probability (of existential threats) and the utility of action (number of life-years at risk). But it should be noted that perverse conclusions are possible when considering the utility value of growing future populations. To avoid such perversity it would be wise to limit considerations to thinking about a stable population continuing into the future, perhaps at New Zealand's current level.

**Figure 2: Projected cumulative number of lives lived in New Zealand for different time periods**

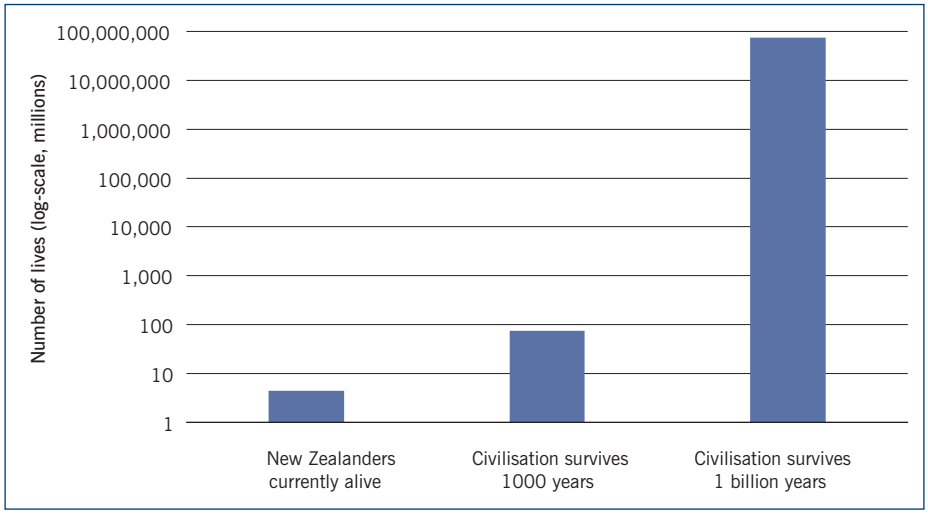
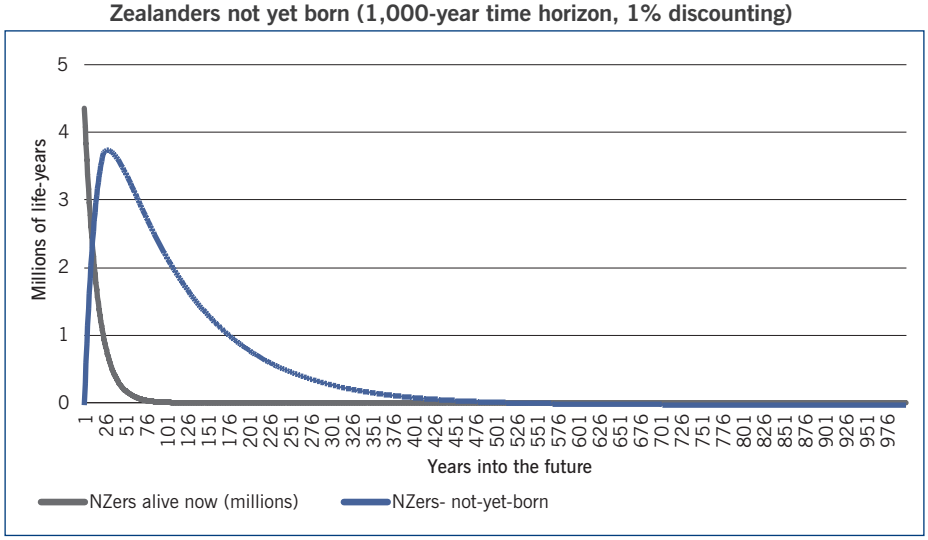



**Figure 3: Annual life-years lived by New Zealanders currently alive and New Zealanders not yet born (1,000-year time horizon, 1% discounting)**




### Example calculations for a possible rational investment in risk reduction

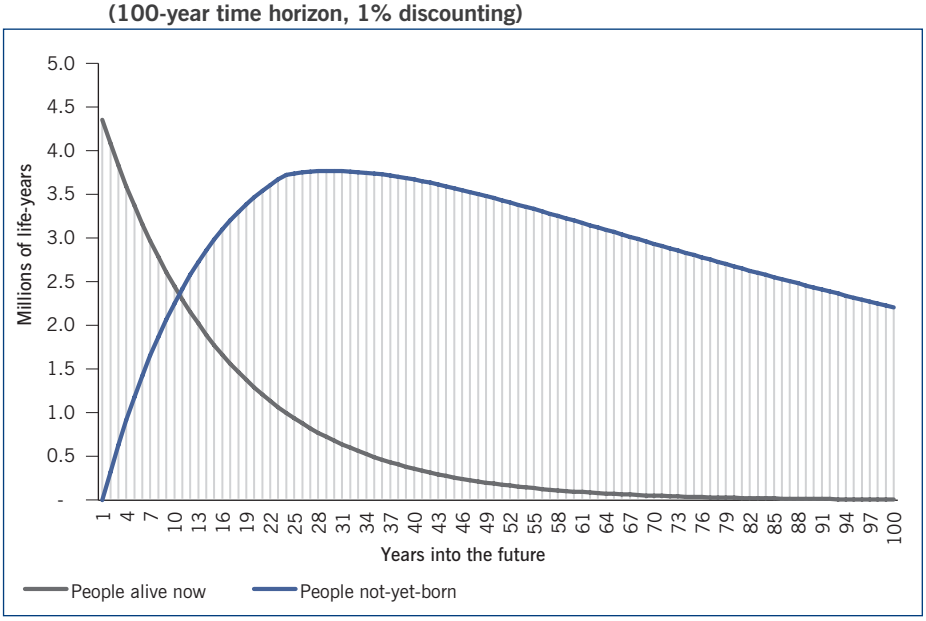
*How many New Zealand lives are at risk?*

Published utilitarian calculations have considered the value of all future lives, whether Earth-bound or dispersed across the universe (Bostrom, 2003, 2013; Jebari, 2014). Here we provide calculations for the number of New Zealand lives at risk under certain assumptions of time horizon (how far in the future lives matter) and discount rate (how much more important are lives now than in the future).

We don't know which utilitarian position the New Zealand public would favour, let alone if utilitarianism itself would be the favoured ethical framework. However, policymakers, under the burden of necessity to act, might accommodate both uncertainty and consilience among value frameworks by applying some moderate discount rate to calculations of future lives, pending the outcome of comprehensive public engagement on what is literally an issue of our very existence.

Figure 2 shows the astronomically large cumulative totals of New Zealand lives that are possible in the future (i.e. around 75,000 billion (75x1012) lives for a stable

**Figure 4: Annual life-years lived by New Zealanders currently alive and not yet born (100-year time horizon, 1% discounting)**




six million population for the expected remaining billion (109) years of Earth being habitable). Even these numbers are potentially miniscule compared to population growth if future New Zealanders join others to become colonists on other planets (Bostrom and Cirkovic, 2008).

Figure 3 shows a view of the next 1,000 years (where we assume a stable population of six million New Zealanders from the year 2040 onwards). Our analysis suggests massive potential numbers: i.e. a cumulative total of 70 million life-years among those already alive (14% of the total) and 515 million life-years among New Zealanders not yet born (at a discount rate of 1%). At what is probably an unreasonably high discount rate of 3%, the total life-years involved in this time period is still 186 million, of which 53 million is among those who are alive now (28% of the total).

A more constrained time scale of just 100 years into the future is one in which some New Zealanders born recently may still be alive throughout and which many of their not-yet-born children will live through (Figure 4). For this period, life-years among the not-yet-born dominate in just ten years' time and comprise 81% of the cumulative 363 million life-years (discounting at 1%).

So, no matter how we calculate it, even conceding that we may care only about presently existing New Zealanders, the numbers of lives and life-years at risk from an existential threat is massive. This is important, because although the probability of an existential threat may be unknown, it is non-zero.

*The probabilities of existential threats*

As a simple exercise, we consider the following: (1) valuing a life-year at per capita GDP (around NZ$45,000 (Kvizhinadze et al., 2015)); (2) the 585 million future New Zealand life-years at risk (70+515 million – see figures above – for a 1,000-year horizon, discounting at 1%); (3) a probability of 0.1% of an existential threat occurring in the next year. Given these values, it would be rational for New Zealand society to invest up to NZ$26 billion in eliminating that risk (though of course by working cooperatively with other countries the cost could be vastly reduced). Yet the probability used in this example may be unrealistically low; some estimates put the risk over the course of the 21st century at 25% or more (Matheny, 2007). Indeed, Lord Martin Rees gives 21st-century human civilisation equal odds (Rees, 2003).

Preventive measures are often thankless investments, because if the disaster fails to befall us, it is often not clear whether it was prevented or simply never eventuated. We need to seriously study these probabilities and mitigation costs (Bostrom, 2013). Investment in the analysis of these risks will allow rational prioritisation.

However, we may never be able to accurately measure the probability of many events (we need to be able to estimate probability, cost of mitigation and utility in order to rank interventions). The theory of 'black swans' (very rare disruptive events) (Taleb, 2007) is a metaphor that describes completely surprising events, with major effect, that can be inappropriately rationalised after the fact. History is full of high-profile, hard-to-predict and rare events that are beyond the realm of normal expectations. Taleb argues that we must build uber-robust or 'antifragile' systems against black swans because we cannot predict them (Taleb, 2012). This might necessitate resilience-style coping measures that are general in nature rather than attempting to prevent specific catastrophes (Jebari, 2014).

Some global catastrophic risks are more likely in the near future than others. Rees has wagered that by 2020 'bioterror or bioerror' will lead to one million casualties in a single event (Kupferschmidt, 2018). The most important countermeasure would be to strengthen our ability to contain such an incident. Nuclear war may also have a significant near-future probability given recent developments (Bulletin of the Atomic Scientists, 2018), while the risk of other threats will probably rise over time – for example, from superintelligent artificial intelligence (Bostrom, 2014).

If we find that the public privilege the value of the future life-years of presently existing people, and discount those of future people, then we find a shifting window of value that moves through time, with a fairly short time horizon (i.e. only ten years using a discount rate of 1%: see Figure 4). It will be existential risks that have the highest probability of occurring in this window which we should probably be most concerned about (perhaps nuclear war). We would then be rational to prioritise such risks according to likelihood and cost of prevention/mitigation. Recent research has attempted to devise novel methods to communicate the level of risk by colour coding in these uncertain settings (Turchin and Denkeberger, 2018).

Once the relevant risks and mitigation strategies (and costs) are identified, we must consider the present opportunity costs of taking action. Preferences in evaluating these costs and the benefits could be grounded in the views obtained from public engagement. We would also

need to consider the present co-benefits of taking action. For example, action to mitigate an existential risk from climate change might reduce the burden of near-future flood damage and other disruptions to agriculture. Ultimately, four factors will drive decision making: the potential impact (including the extent that the risk may really be existential); the probability of occurrence; the capacity to reduce the risk; and the cost of risk reduction. All public expenditure has opportunity costs, and ideally the different risk mitigation strategies will be evaluated for relative cost-effectiveness. Even so, some may be cost-saving (for example, removing government subsidies to the oil and gas exploration industry as one component of preventing further climate change).

New Zealand is a small country, but we can contribute to global knowledge about how to define, approach and prepare for existential threats. New Zealand has previously campaigned for nuclear arms control and could work with likeminded countries to strengthen action against climate change. New Zealand has had successes in terms of governments looking to the longer term, including the New Zealand Superannuation Fund, Earthquake Commission and Children's Commissioner, but we could go further and strengthen future-oriented commitments (Boston, 2017). Once we know what New Zealanders think, we can engage on the international stage to build resilience.

**Conclusion**

No matter how the number of future lives and life-years is calculated, the result is that gargantuan numbers of currently living and, especially, future people are potentially threatened by existential risks. Policymakers should therefore give more consideration to the future and preventing such existential risks. Of all the risks to things we value, some are urgent and some are important, and we need to focus on those that are both urgent and important (Bostrom, 2014, p.256). A just process for resource allocation demands that we consider future generations but also account for solidarity with the present. We need to establish what New Zealand society wants and values. We need to know what people think about the future life-years of people alive now and those not yet born. The philosophical attitude towards future people that a global community takes will determine the kinds of utility calculations that are required. There are threats that demand action now, such as nuclear war, and, as we move forward, understanding of our values will inform appropriate policy to rationally and optimally address other existential risks.

1 Matthew Boyd conducted literature searches, interpreted the data, wrote the manuscript and contributed important philosophical content. Nick Wilson conceived the idea, performed data analysis and interpretation, and contributed important intellectual content at manuscript drafting stage. The authors declare that they have no competing interests.

**References**


Adler, M. (2009) 'Future generations: a prioritarian view', *George Washington Law Review*, 77, pp.1478–520

Arrhenius, G. (2000) 'An impossibility theorem for welfarist axiologies', *Economics and Philosophy*, 16, pp.247–66

Arrhenius, G. and W. Rabinowicz (2010) 'Better to be than not to be?', in H. Joas and B. Klein (eds), *The Benefit of Broad Horizons: intellectual and institutional preconditions for a global social science*, Leiden: Brill

Boston, J. (2017) *Safeguarding the Future: governing in an uncertain world*, Wellington: Bridget Williams Books

Bostrom, N. (2003) 'Astronomical waste: the opportunity cost of delayed technological development', *Utilitas*, 15 (3), pp.308–14

Bostrom, N. (2013) 'Existential risk prevention as global priority', *Global Policy*, 4 (1), pp.15–31

Bostrom, N. (2014) *Superintelligence: path, dangers, strategies*, Oxford: Oxford University Press

Bostrom, N. and M. Cirkovic (eds) (2008) *Global Catastrophic Risks*, Oxford: Oxford University Press

Boyd, M. and N. Wilson (2017) 'Rapid developments in artificial intelligence: how might the New Zealand government respond?', *Policy Quarterly*, 13 (4), pp.36–43

Bromell, D. (2012) 'Doing the right thing: ethical dilemmas in public policy making', working paper, Centre for Theology and Public Issues, University of Otago

Broome, J. (2005) 'Should we value population?', *Journal of Political Philosophy*, 13 (4), pp.399–413

Bulletin of the Atomic Scientists (2018) 'Doomsday Clock: timeline', retrieved 30 January from https://thebulletin.org/timeline

Council of the New Zealand Ecological Society (1985) 'The environmental consequences to New Zealand of nuclear war in the northern hemisphere', *New Zealand Journal of Ecology*, 8, pp.163–74

Daniels, N. and J. Sabin (1997) 'Limits to health care: fair procedures, democratic deliberation and the legitimacy problem for insurers', *Philosophy and Public Affairs*, 26, pp.303–50

Deery, O., T. Davis and J. Carey (2014) 'The free-will intuitions scale and the question of natural compatibilism', *Philosophical Psychology*, 28 (6), pp.776–801, doi: 10.1080/09515089.2014.893868

Freiberg, A. and W. Carson (2010) 'The limits to evidence-based policy: evidence, emotion and criminal justice', *Australian Journal of Public Administration*, 69 (2), pp.152–64

Gluckman, P. (2011) *Towards Better Use of Evidence in Policy Formation: a discussion paper*, Auckland: Office of the Prime Minister's Science Advisory Committee

Gosseries, A. and L. Meyer (eds) (2009) *Intergenerational Justice*, Oxford: Oxford University Press

Gruen, G., D. Kelly and S. Gorecki (2011) 'Wellbeing, living standards, and their distribution', paper presented as part of the New Zealand Treasury academic guest lecture series

Heyd, D. (2008) 'A value or an obligation? Rawls on justice to future generations', in A. Gosseries and L. Meyer (eds), *Intergenerational Justice*, Oxford: Oxford University Press

Hope, T. (2001) 'Rationing and life-saving treatments: should identifiable patients have higher priority?', *Journal of Medical Ethics*, 27, pp.179–85

Jebari, K. (2014) 'Existential risks: exploring a robust risk reduction strategy', *Science and Engineering Ethics*, 21 (3), pp.541-54, doi: 10.1007/s11948-014-9559–3

Kahane, G. (2013) 'The armchair and the trolley: an argument for experimental ethics', *Philosophical Studies*, 162, pp.421–45

Kupferschmidt, K. (2018) 'Taming the monsters of tomorrow', *Science*, 11 January

Kvizhinadze, G., N. Wilson, N. Nair, M. McLeod and T. Blakley (2015) 'How much can society spend on life-saving interventions at different ages while remaining cost effective? Estimates using New Zealand health system costs, morbidity, and mortality data', *Population Health Metrics*, 13 (15)

Matheny, J.G. (2007) 'Reducing the risk of human extinction', *Risk Analysis*, 27 (5), pp.1335–44, doi: 10.1111/j.1539-6924.2007.00960.x

Meyer, K. (2018) 'The claims of future persons', *Erkenntnis*, 83 (1), pp.43–59, doi: 10.1007/s10670-016-9871-1

Nadelhoffer, T., J. Shepard, E. Nahmias, C. Sripada and L. Ross (2014) 'The free will inventory: measuring beliefs about agency and responsibility', *Consciousness and Cognition*, 25, pp.27–41

Narveson, J. (1967) 'Utilitarianism and new generations', *Mind,* 76, pp.62–72

New Zealand Treasury (2017) *CBAx Tool User Guidance: guide for departments and agencies using Treasury's CBAx tool for cost benefit analysis*, Wellington: New Zealand Government

Norton, B. (1989) 'Intergenerational equity and environmental decisions: a model using Rawls' veil of ignorance', *Ecological Economics*, 1 (2), pp.137–59

Nussbaum, M. (2000) *Women and Human Development: the capabilities approach*, Cambridge: Cambridge University Press

Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press

Rawls, J. (1971) *A Theory of Justice*, Oxford: Oxford University Press

Rawls, J. (1987) 'The idea of an overlapping consensus', *Oxford Journal of Legal Studies*, 7 (1), pp.1–25

Rees, M. (2003) *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

Rockstrom, J., W. Steffen, K. Noone, A. Persson, F. Chapin, E. Lambin et al. (2009) 'Planetary boundaries: exploring the safe operating space for humanity', *Ecology and Society*, 14 (2)

Scheffler, S. (2013) *Death and the Afterlife*, Oxford: Oxford University Press

Singer, P. (1972) 'Famine, affluence and morality', *Philosophy and Public Affairs*, 1 (3), pp.229–43

Sonderholm, J. (2013) 'World poverty, positive duties, and the overdemandingness objection', *Politics, Philosophy and Economics*, 12 (3), pp.308–27

Taleb, N. (2007) *The Black Swan*, New York: Random House

Taleb, N. (2012) *Antifragile: how to live in a world we don't understand*, New York: Random House

Tanyi, A. and M. Bruder (2014) 'Consequentialism and its demands: a representative study', *Journal of Value Inquiry*, 48 (2), doi: 10.1007/s10790-014-9418-0

Tarsney, C. (2017) 'Does a discount rate measure the costs of climate change?', *Economics and Philosophy*, 33 (3), pp.337–65

Tonn, B. and D. Stiefel (2014) 'Human extinction risk and uncertainty: assessing conditions for action', *Futures*, 63, pp.134–44

Turchin, A. and D. Denkeberger (2018) 'Global catastrophic and existential risks communication scale', *Futures*, forthcoming

Weale, A. (1979) 'Statistical lives and the principle of maximum benefit', *Journal of Medical Ethics*, 5, pp.185–95

Weitzman, M. (1998) 'Why the far-distant future should be discounted at its lowest possible rate', *Journal of Environmental Economics and Management*, 36, pp.201-8

World Economic Forum (2014) *Global Risks 2014*, Geneva: World Economic Forum

World Economic Forum (2017) *The Global Risks Report 2017*, Geneva: World Economic Forum