### Securing Political Accountability to Future Generations with Retrospective Accountability[[1]](#footnote-0)

Keywords: Future generations, political accountability, democracy, mechanism design, institutional design, game theory, short-termism

Abstract: Political short-termism costs the global economy hundreds of billions to trillions of dollars annually, and leads to many millions of deaths from disasters and suboptimal spending. In this paper, I propose a futures assembly explicitly incentivised to promote the interests of future generations as a promising strategy to ameliorate short-termism. The assembly I propose is governed by citizens randomly selected from among the populace, who are rewarded in the future to the extent that they successfully promote the welfare of future generations. The most initially promising such retrospective accountability mechanism is an iterated mechanism, in which each generation of policymakers rewards the previous generation for their policy choices and for their own evaluation of the previous generation. Such an assembly, supplemented with the duties, powers, and tools that I describe, and with appropriate support from archivists and experts, may manage to represent generations many years into the future.

A number of recent writers have argued that the obligations of modern states to future generations may far outstrip their obligations to their present citizens, given the vast number of people who will exist in the future and whose livelihoods depend on our actions (Beckstead 2013, Greaves and MacAskill 2019, Tarsney 2019). And yet modern states do precious little on behalf of future generations, opting to allow and incentivise destructive practices such as the widespread burning of fossil fuels, while failing to take preventative measures that could deter global pandemics and other catastrophes.

 The state is plagued with problems of political short-termism: the excessive priority given to near-term benefits at the cost of future ones (González-Ricoy and Gosseries 2016B). By the accounts of many political scientists and economists, political leaders rarely look beyond the next 2–5 years and into the problems of the next decade. There are many reasons for this, from time preference (Frederick et al. 2002, Jacobs and Matthews 2012) to cognitive bias (Caney 2016, Johnson and Levin 2009, Weber 2006) to re-election incentives (Arnold 1990, Binder 2006, Mayhew 1974, Tufte 1978),[[2]](#footnote-1) but all involve foregoing costly action in the short term (e.g. increasing taxes, cutting benefits, imposing regulatory burdens) that would have larger moderate- to long-run benefits. Such behaviour is detrimental not only to the generations of people who are to come, but also to the large number of existing citizens who still have much of their lives left to lead.

 A small literature in philosophy, economics, and political science considers what modern governments might do to ameliorate short-termism and incentivise states to adequately prioritise the interests of future generations in policymaking (see especially Caney 2016, John and MacAskill 2021, and González-Ricoy and Gosseries 2016C). One core unsolved problem in this literature is the problem of developing incentive mechanisms for present policymakers to promote the welfare of future generations. For most disenfranchised groups, simply giving them the vote and ensuring a functional democratic political system ensures that their preferences are represented in government. But future generations cannot cast a ballot—so how can we ensure that governments adequately protect them?

I have previously criticised extant attempts to solve this problem (John 2020, John 2023). This chapter is my attempt to develop a successful strategy for incentivising the present generation to act on behalf of future generations: building a novel *futures assembly* which is explicitly incentivised to promote the interests of future generations. The assembly I propose is governed by citizens randomly selected from among the populace, who are rewarded in the future to the extent that they are successful in promoting the welfare of future generations. In Section 1, I outline a novel futures assembly and detail the two essential features of its mechanism design: sortition (random selection of assembly members) and retrospective liability (rewarding assembly members in the future, based on their past performance vis-a-vis safeguarding the future). Section 2 elaborates the function of the assembly, including the duties and powers that it should have. The assembly’s structure is spelled out in Section 3, including the constitution of the assembly, the role of experts, and its research and media departments. Section 4 briefly discusses how the assembly should interact with the rest of government. I conclude with some reflections on the epistemic status of the institution I have proposed. In particular, I consider the extent to which this mechanism can extend today’s time horizons—whether it can plausibly improve the world in not merely 1,000 years, but also one million years hence.

#### 1. A Forward-looking Assembly

Two key questions for any proposed political institution are: (1) Who serves? and (2) What are their incentives? The body I propose is populated by random selection of political decision-makers from among the populace, who are educated on issues affecting future generations prior to making political decisions, to ensure that the people who serve are unelected, informed, non-polarised, and representative. It is incentivised by retrospective accountability, a mechanism on which decision-makers are rewarded years after their tenure based on the effects of their decisions over that period, to ensure that the people who serve are incentivised to promote the well-being of future generations. In this section I articulate and defend these two core features of the proposed representative mechanism.

#### 1.1. Sortition

“Citizens assemblies” have been employed for consultation and information-gathering purposes throughout the world. These randomly-selected groups of citizens provide deliberative and non-binding advice to the government in consultation with recognized experts. One of the most high-profile initiatives was Ireland’s 100-member Citizens’ Assembly, which was established in 2016 and tasked with considering questions related to abortion, fixed-term parliaments, referenda, population aging, climate change, and gender equality. The deliberations of the Irish assembly provoked a referendum to remove Ireland’s constitutional ban on abortion and substantially shaped Ireland’s Climate Action Plan (Coleman et al. 2019).

The success of the Irish assembly and of citizens’ assemblies around the world reveals the promise of citizens assemblies tasked with the explicit mandate to represent future generations, or “futures assemblies.” A general futures assembly, constituted by a stratified random sample of the general population, would have numerous features that predict success at combating short-termism (John and MacAskill 2021). Being an unelected and publicly-funded body, a futures assembly would be insulated from the election and fundraising incentives that pressure elected officials to focus on near-term, visible issues. Being randomly selected, it would be statistically representative of the general population. And citizens’ assemblies have a demonstrated aptitude in “laboratory” and real-world experiments for reducing the deleterious effects of partisanship on careful, long-term deliberation (Fishkin and Luskin 2005, Fishkin et al. 2017, List et al. 2013). In the most recent major assembly, the Climate Assembly UK, 98% of assembly members claimed to have understood almost everything that those in their deliberation groups had said, and 94% felt respected by their fellow participants under disagreement (with none feeling disrespected) (Climate Assembly UK 2020). Finally, citizens’ assemblies are more informed than ordinary voters due to their deliberations with experts, reducing the deleterious effect of policy uncertainty on short-term policy support (cf. Jacobs and Matthews 2012).

A general futures assembly may need no incentive to reflect carefully on the interests of future generations beyond an explicit mandate to do so. Some limited evidence from the Kochi University Research Institute for Future Design suggests that when parents are explicitly asked to cast votes on behalf of their children, they vote for different parties than they normally would vote for in a sizable minority of cases (Aoki and Vaithianathan 2012). This is a promising sign that those who are asked explicitly to represent other generations do not simply use the opportunity to promote their own agenda, but rather aim to promote the interests of the relevant generation, and thereby adopt longer time horizons for political decision-making. This is further supported by evidence that actors within institutions tend to be compelled to follow norms that are consistent with the established culture of their institution (Goodin 1986, MacKenzie 2016, Steiner et al. 2005: 127).

#### 1.2. Retrospective Accountability

An explicit mandate may be sufficient to motivate futures assembly members to give substantial consideration to future generations in their recommendations. But there are a number of reasons why their consideration of future generations may be inadequate, including public pressure, value misalignment, value drift, weakness of will, intrinsic time discounting (or impatience), and corruption. Moreover, evidence that political actors follow the established rules of their institutions is limited, and we do not know the universality of its scope or the strength of its implications. We can have much greater confidence in the decisions of a randomly-selected futures assembly if it is directly incentivised to promote the welfare of even distant future generations.

An underexplored mechanism for aligning incentives with the interests of future generations involves *retrospective accountability*. The central problem of representing the interests of future generations in government is that of making political actors accountable to future generations. Future generations cannot vote in our elections, nor can they sanction or protest the decisions of their forebears. Retrospective accountability solves the accountability problem by rewarding policy-makers years into the future in proportion to the effects of their policy on the long run. A simple mechanism of retrospective accountability would involve empowering a body of future auditors—say, 30 years from now—to decide on the pension bonus of the decision-makers today based on how successfully these decision-makers promote the interests of future people. This would provide decision-makers today with a positive financial incentive to look to the future—at least 30 years from now—when making any decisions. Such a mechanism would yield a significant advance on the time horizons of present institutions.[[3]](#footnote-2)

There are a variety of retrospective accountability mechanisms that can be employed by governments. First, governments can place an age limit on the relevant class of political decision-makers and reward people late in life. If the age limit is 30-40 years, and they are rewarded around the age of 60, this could extend political time horizons by 20-30 years. Second, governments can choose decision-makers who are parents or otherwise connected to children, and reward these children later in life based on the decisions of the parents. This could extend political time horizons by 40 years, or even longer. Third, governments can find other rewards that incentivise policymakers even when the rewards are given out after they have died, such as by promising to further their projects, build their communities, or improve their legacy. This could in principle extend the time horizons of policymakers indefinitely (discounted by the likelihood that these commitments will be fulfilled), but we do not at this time know how strongly such promises could motivate policymakers, or which such mechanisms are most powerful. Fourth, governments can employ retrospective accountability mechanisms *iteratively*, selecting a sequence of policymakers who each decide on the later bonus of a previous generation, with each of their own future bonuses tied to the opinions of the following generation. I think that such an iterated accountability mechanism is promising, and it warrants detailed analysis.

#### 1.3. Iterated Retrospective Accountability

One possible approach would exploit strategic iteration of this mechanism to extend the time horizons of government far into the future. On the iterated variant, the future auditors who decide on the future bonuses of present decision-makers *themselves* face a financial incentive to look again into the future. For their own financial situation will be tied to the evaluations of the *next* generation of auditors, who will determine their pension bonuses. To get a nice retirement bonus, future auditors have an incentive to evaluate present decision-makers in accordance with the preferences of the next generation of auditors, and so present decision-makers have an incentive to satisfy the preferences of the auditors two generations—60 years—from now. And so iterated, until we have extended the horizons of government to the longest time period relevant for political decision-making. On the simplest implementation of such accountability measures, the assemblies are also the auditing bodies: each assembly decides on the bonus of the assembly 30 years prior.

 We should give special attention to two subspecies of iterated accountability mechanisms, differentiated by whether there is a known period when the iterated mechanism terminates. In the “Final Tribunal of Justice” model, there is a known final auditor in the series whose judgments are not incentivised by a subsequent auditor’s retrospective rewards. On the “Infinite Justice” model, there is no known final auditor; the series continues indefinitely, with some positive probability of termination at every period, due to either repeal or social collapse. In this section I show that both subspecies of iterated accountability mechanism can succeed at incentivising political decision-makers to significantly extend their time horizons, under certain conditions.

#### 1.3.1. The Final Tribunal of Justice

In the Final Tribunal of Justice model, the iterated accountability mechanism has a known final auditor. This auditor is not evaluated by any further auditors, and so has no incentive from the mechanism to cooperate with the rules of the scheme: namely, to reward the previous generation to the extent that they chose the “optimal strategy,” that is, chose the optimal policies and adopted a reward strategy that optimally incentivised the previous generation to choose optimal policies and rewards.

For this reason, the Final Tribunal of Justice succeeds in incentivising the chain of auditors to choose the optimal strategy if and only if each generation *n* is sufficiently confident that generation *n + 1* is sufficiently confident that generation *n + 2* is sufficiently confident that … generation *n + i*, the final generation, will be motivated to choose the optimal strategy.[[4]](#footnote-3) The precise degree of confidence required depends on the auditors’ own preferences about whether to cooperate or defect, and the size of the reward for cooperation.[[5]](#footnote-4) For ease of exposition, I’ll assume that the degree of confidence required by each generation is full belief.

Suppose that a generation *n* believes that generation *n + 1* believes that generation *n + 2* believes that … generation *n + i*, the final generation, will be motivated to choose the optimal strategy. Then generation *n* can use backwards induction to see that they, too, ought to use the optimal strategy. Using the optimal strategy involves rewarding the previous generation *iff* they chose the optimal strategy. So, if the final auditor *n* chooses the optimal strategy, they will reward the previous generation *n - 1* *iff* *n - 1* chooses the optimal strategy. If *n - 1* believes this, they will be motivated to choose the optimal strategy. And if *n - 2* believes that *n - 1* will be motivated to choose the optimal strategy, they also believe that *n - 1* will be motivated to reward the previous generation *n - 2* *iff* *n - 2* chooses the optimal strategy. So *n - 2* ought to infer, via backwards induction, that they will be rewarded if they choose the optimal strategy. And so iterated, until we reach the first auditor, who too can infer, via backwards induction, that they will be rewarded if they choose the optimal strategy.

By contrast, suppose that some generation *n* does *not* believe that generation *n + 1* believes that generation *n + 2* believes that … generation *n + i*, the final generation, will be motivated to choose the optimal strategy. Then *n* cannot infer that *n + 1* will reward *n* for choosing the optimal strategy. In fact, if *n* is sufficiently confident that it is *not* the case that generation *n + 1* believes that generation *n + 2* believes that … generation *n + i*, the final generation, will be motivated to choose the optimal strategy, then *n* can employ backwards induction to see that they will *not* be rewarded for choosing the optimal strategy. Consider a random generation *n + k* such that *n* is confident that generation *n + 1* believes that generation *n + 2* believes that … generation *n + k* does *not* believe that the subsequent generation will be motivated to choose the optimal strategy. Then *n* can infer that *n + 1* can infer that *n + 2* can infer that … *n + k - 1* can infer that *n + k* will have no incentive to reward *n + k - 1* for choosing the optimal strategy. And so *n* can infer that *n + 1* can infer that *n + 2* can infer that … *n + k - 1* will have no incentive to choose the optimal strategy. And so *n* can infer that *n + 1* can infer that *n + 2* can infer that … *n + k - 1* will have no incentive to reward *n + k - 2* for choosing the optimal strategy. And so on, finally licensing the inference that *n + 1* has no incentive to choose the optimal strategy, providing no incentive, in turn, to *n* to choose the optimal strategy.

So, the Final Tribunal of Justice succeeds in incentivising the chain of auditors to choose the optimal strategy if and only if each generation *n* is sufficiently confident that generation *n + 1* is sufficiently confident that generation *n + 2* is sufficiently confident that … generation *n + i*, the final generation, will be motivated to choose the optimal strategy. This condition would be trivially met if there were common knowledge that the final generation will be motivated to choose the optimal strategy. So one way to ensure cooperation is to make it public that the final generation *is* motivated to choose the optimal strategy. This could happen if the scheme were committed to the choice of a final generation with the right motivational profile—perhaps people who are deeply motivated to cooperate with the scheme, or people who are motivated by justice for their own generation—and making it public that this is the case. In some light this does not seem particularly difficult. By default, it would be surprising if the final generation were *not* significantly motivated to judge the previous generation harshly if the previous generation had failed them, and to judge them favourably if the previous generation had helped them. Juries have no external motivation to judge court cases aptly, but their sense of justice and the rules of the scheme ensure that juries work reasonably well. However, if it were unclear whether the final auditor would be properly motivated, the scheme would fall apart.

#### 1.3.2. Infinite Justice

On the Infinite Justice model, the iterated accountability mechanism has no known final auditor. So each auditor in the series has incentive from the mechanism to cooperate with the rules of the scheme and to reward the previous generation to the extent that they chose the optimal strategy. In the Final Tribunal of Justice model, each generation must believe that the last generation will be intrinsically motivated to follow the scheme. But in the Infinite Justice model no such guarantee is required.

 The Infinite Justice model assumes that policymakers are motivated to choose short-termist policy. In the absence of intervention, then, there is one unique perfect subgame equilibrium: policymakers will always choose the short-termist policy and cannot rationally deviate from the choice of a short-termist policy. By offering a bonus to policymakers that depends on their policy choice, the Infinite Justice model allows for multiple perfect subgame equilibria depending on their expectations about what future auditors will reward. If a generation of auditors expects future auditors to reward policy with time horizon H, and they are rational, then this generation of auditors will in fact choose policy with time horizon H, under the specific conditions outlined in 1.3.3.

 The Infinite Justice model can sustain cooperation on ideal policy because, so long as the per-generation probability that the mechanism will collapse is sufficiently low, the auditors are unable to determine which generation is the last generation. No auditor can be confident that they or the next generation will be the last generation, and so they cannot be confident that they or the next generation have no incentive to cooperate with the scheme. As long as there is a sufficiently high chance that there will be two subsequent generations, an auditor can expect to reap rewards for choosing the optimal policy.

The Appendix includes a proof from economist Charlotte Siegmann demonstrating that auditors will rationally be compelled to choose the policy they expect future auditors to reward rather than choosing the short-termist policy precisely under these conditions.

 So it is provable that the Infinite Justice model can sustain cooperation on a set of policy goals across many generations. But it is compatible with multiple perfect subgame equilibria. A corrupt line of auditors could collectively coordinate to give the previous generation a bonus no matter what. As long as the auditors expect the next few generations to play the same game, rewarding the previous auditors no matter what, they could continue to adopt the short-termist policies they prefer. So we need to have reason to expect that they will coordinate on *long*termist policy goals, such that most generations of auditors expect future auditors to reward them for adopting a long time horizon. In particular, we need to have reason to expect that longtermist policy is a salient Schelling Point on which each generation can coordinate.

 In the real world, we should expect an equilibrium to emerge around long rather than short time horizons. First, as discussed previously, political actors find following the mandates of their office intrinsically rewarding, and this office will mandate adopting long time horizons. This provides direct incentive for policymakers to choose a long time horizon, but it also gives each generation reason to expect that future generations of policymakers will also choose a long time horizon, and so it gives them reason to expect that they will be rewarded for doing the same. Second, only a cooperative equilibrium with long time horizons will be communicated to policymakers, and so by default we should expect that they will settle on this equilibrium. And third, in any sensible state, a corrupt and persistent scheme of defection from the purposes of the assembly will be repealed. And given that the bonus of each generation depends on the system *not* being repealed before the audit, a rational futures assembly will choose not to defect, given that doing so would eliminate their reward. That is, though the rational choice model assumes that the probability that the futures assembly will be repealed is independent of choices made by the assembly, this is clearly not true. If the first assembly makes a series of short-termist policy decisions and the second assembly rewards them for it, the political contemporaries of the second assembly will repeal the assembly for failing to abide by the rules of the office, and for failing to promote their interests. So a rational such assembly will not make a series of short-termist policy decisions.

 The arational pressures towards intergenerational cooperation due to intrinsic motivation to follow the norm and the narrative set by the office, as well as external pressures of repeal against a corrupt office, should imply a cooperative equilibrium around long time horizons, even where participants in the scheme deviate from game-theoretic rationality. Without small-scale intergenerational cooperation experiments it is difficult to know how successful a scheme of Infinite Justice would be in practice. But the underpinning game theory illuminates the model as a promising mechanism for further investigation and highlights the institutional features that are relevant to such a system working in the real world, as I next discuss.

#### 1.3.3. Conditions of Success

Mechanisms using either the Final Tribunal of Justice model or the Infinite Justice model need to satisfy some conditions if they are to succeed. These conditions are modeled explicitly on the Infinite Justice rational choice model found in the Appendix, captured by the formula 𝛾 < (1 - 𝛿). On this model, the expected value of the bonus to the auditors—discounting for the chance that they die or the next generation defects or the next generation knows when the mechanism will be repealed—is higher than the expected intrinsic reward of the next generation’s giving them the bonus. If, in either model, this condition fails to hold, then the mechanism cannot sustain cooperation.

For the condition to hold, a few things must happen. First, and roughly, auditors cannot be nearly as excited to give money to the previous generation as they are to receive money themselves. That is, they need to value their own money more than they value the previous generation’s money, at least by several times. Fortunately, it is generally true that people value their own money at least several times more than they value strangers’ money. Second, auditors cannot have too high an expectation of dying before they receive and can benefit from the money (or they must value passing it on to their kin about as much as they value receiving it). So the temporal distance between auditing and audited generations cannot be too far. Third, auditors cannot be too confident that the office will be repealed before the next auditing cycle, or that the next auditor will know when the office will be repealed. For this reason, institutional longevity should be a key desideratum of retrospective accountability mechanisms. And fourth, auditors need to have sufficiently high confidence that they can do their job well enough to receive a bonus. This means the conditions for the bonus cannot be overly demanding, and that futures assembly members need tools to be confident that they can do their jobs well.

This final criterion of success entails some further knowledge conditions. In particular, a majority of auditors must be sufficiently confident that the next generation will cooperate, and so they must be sufficiently confident that the next generation understands the optimal strategy profile and the reasons for cooperation. If they have too high an expectation of ignorance, they will place a low expected value on cooperation, since the link between their cooperation and receiving their bonus will be tenuous.

 In this section I’ve proposed a bare-bones futures assembly that employs two core design features: sortition and retrospective accountability. Sortition functions to select a statistically representative unelected body of citizens to serve in the assembly. Retrospective accountability functions to solve a kind of intergenerational principal-agent problem, making the incentives of present decision-makers tied to the choices of future generations. The discussion has raised two key desiderata for a successful futures assembly employing these design features:

(1) T**he Epistemological Criterion**: futures assemblies must have tools for solving difficult epistemological problems in order to have the confidence that they can do their jobs well enough to be rewarded, and so that later assemblies will have the capacity to correctly evaluate them.

(2) T**he Longevity Criterion**: futures assemblies need to be highly resilient, with sufficiently high institutional longevity to give each policymaker confidence that the institution will persist during the period under which they are rewarded (or several subsequent periods on the iterated model).

The next three sections put some muscle on the proposed assembly’s bones. The features that help the futures assemblies to satisfy the Epistemological Criterion and the Longevity Criterion will be woven throughout the following three sections, which focus on the assembly’s job, powers, and tools; its structure; and its relationship with the rest of government, respectively.[[6]](#footnote-5)

#### 2. Duties, Powers, and Tools

At a very general level, the Futures Assembly is tasked with and incentivised to represent future generations. But providing the assembly with more specific, concrete duties will help to ensure that they are motivated, and have appropriate guidance, to pursue the right goals. What duties should national and international governments give to an assembly with the design features I’ve described, and with the functional role of representing the interests of future generations?

 The two central duties of the office are to (i) exercise its powers and work with government to produce policy that promotes the welfare of future generations, and (ii) evaluate the decisions of one previous assembly and determine a reward for its members, as a function of their success in their duties. In the iterated retrospective accountability model, the reward is a function of the previous assembly’s performance on (i) and (ii). On the standard retrospective accountability model, the reward is a function of the previous assembly’s performance on only (i). As discussed, this reward may be a monetary reward for the previous assembly, or it may be a different sort of reward, such as a gift to their descendants, the continued promotions of their projects, or the protection of their legacy.

#### 2.1. Policy Duties

Let’s consider the first duty of the assembly: to exercise its powers and work with the government to produce policy that promotes the welfare of future generations. Two parts of this duty require further elaboration and precissification. The first is the powers and tools that the assembly can use to achieve its goal. The second is what, precisely, it means for assemblies to have the goal of promoting the welfare of future generations.

 A central set of tools that any futures assembly will have at its disposal are the government-facing “soft powers” of persuasion, research and investigation, and advice-giving, as well as communication with the general public. More specifically, assemblies have the powers of a direct line to legislators, the ability to set their own research and policy agendas outside the reaches of the political business cycle, and the power to call on experts and convene summits. They also have high institutional legitimacy and status and a PR team to share ideas and craft public narratives to inspire movements among the public.

Despite not having any teeth, even purely soft-power political institutions often have a very significant effect on government decision-making. Many citizens’ assemblies have this profile, such as the aforementioned Irish Citizens’ Assembly, which provoked a referendum to remove Ireland’s constitutional ban on abortion and substantially shaped Ireland’s Climate Action Plan. In-government research institutes such as the Office of Technology Assessment (OTA) have no formal powers at all. Yet a 1990 study by the Carnegie Commission on Science, Technology, and Government found that OTA reports were “useful” to “very useful” to 91% of congressional staff (Bimber 1990). One analysis found that the OTA’s 1980s studies on synthetic fuels “helped secure approximately $60 billion in savings” (Tudor and Warner 2019).

 Recent research suggests that another soft-power institution, the European ombudsman, also has a significant effect on government policy (Beckman and Uggla 2016, Finkel 2006, Koo and Ramirez 2009: 1330, Reif 2011: 286), and even that it played a key role in the Democratisation of European Communist countries (Gilligan 2010: 578). For example, Baranovsky (2016) finds that “in a fixed effects model, both the length of service of the [Russian Ombudsman for Human Rights] and the relative busyness of the office (number of complaints per capita) have had an effect on a region’s corruption index.”

 The second set of powers to review the structure, staff, and rules of the futures assembly are similarly “soft”, in that they do not have the power to coerce or compel any other government institution. This set of powers is key to ensuring the relevance of the institution for many generations to come.[[7]](#footnote-6)

 These two powers of persuasion and of self-review make up the core tools that a soft-power futures assembly can use to promote its policy agenda. An institution with just these powers will likely do best by the lights of the Longevity Criterion. A major reason that future-oriented institutions are repealed today is that they have too much formal power (Jones et al. 2018). So the future-oriented institution that is least likely to be repealed would plausibly have very minimal formal powers.[[8]](#footnote-7)

We’ve considered the first part of the futures assembly’s duty to exercise its powers and work with the government to produce policy that promotes the welfare of future generations: the powers and tools that the assembly can use to achieve its goal. Now we can consider the second part: what, precisely, it means for assemblies to have the duty of promoting the welfare of future generations. Here are three key questions: (1) what normative framework should assemblies apply to evaluate welfare changes in a generation?, (2) on what time horizon should the assemblies operate?, and (3) what issues should the assemblies prioritise?

 The first question is perhaps the most consequential for what the assembly’s goals should be, but it is not one that I can answer here. What particular philosophical theory of well-being futures assemblies should adopt, and the broader question of how they rank states of affairs with respect to value, is, in modern democracies, a matter for public deliberation, and must ultimately be beholden to the deliberations of each assembly, to allow for these assemblies’ self-government and for changes in the citizens’ terminal goals as our moral understanding advances and our preferences change.

 It is tempting to answer the second question in a similar way. However, there are several reasons why the rough time horizons of the futures assemblies should be determined by prior legislation. The first is that the time horizons of futures assemblies must be consonant with their institutional incentives. If, for example, an assembly is required to promote the interests of people more than one century from now, but is rewarded 30 years into the future, then it will be poorly incentivised to achieve its own mandates, and indeed will face the dualing, contradictory incentives of meritocratic reward and intrinsic motivation to act on the prescribed mandates. In addition to weakening their incentives to do any particular thing, this could cause confusion about goals and internal divisions within the assembly. The second reason is the need for temporal coordination across the various assemblies, to ensure that all generations are protected. To see how this could be a problem, imagine that the assemblies in power from 2030–2059 optimise for 30 year time horizons, whereas the assemblies in power from 2060 optimise for 60 year time horizons. In such a case, no futures assembly is responsible for protecting the people who live from the years 2090–2120. If assemblies all adopt the same time horizons, we can ensure the protection of all generations. With that said, our ability to predict the future changes with time. Sometimes forcasting becomes easier due to better tools and techniques, and sometimes it becomes worse due to entrance into periods of explosive growth and change. And if only for this reason, time horizons should be subject to adaptation through the futures assemblies’ intentionally sluggish powers of self-review, alongside corresponding changes to the time horizons of the institutions’ incentives mechanisms, if two successive assemblies agree to such changes.

 On to the third question: what issues should assemblies prioritise? At a very basic level, it is a fairly straightforward matter that futures assemblies should consider all areas of policy that concern future generations on the relevant time horizon. This is because decisions in one policy domain that affects future generations (such as green energy) very often have effects in another policy domain that affects future generations (such as the workforce or economic growth). To have balanced policy that attends to all of the needs of future generations, and does not attend to one need at the expense of others, any sole institution charged with representing future generations must consider all areas of policy that concern future generations.[[9]](#footnote-8)

#### 2.2. Evaluation Duties

With the first set of duties precissified, and the powers to execute these duties elaborated, we can now turn to the assemblies’ second set of duties: to evaluate the decisions of one previous assembly and determine a reward for its members, as a function of their success in their duties.

While I have saved the disscusion of evaluation duties for last, these are the first duties that a futures assembly should work to fulfill. Futures assemblies are constituted by ordinary citizens, so they have a lot to learn before they can begin engaging in successful policymaking. The exercise of a futures assembly’s evaluation duties are part of its early learning phase—a period of about one year wherein assembly members consult with experts on issues of long-term importance; evaluate the work of a previous assembly; and attend seminars about civics, the legal structure and history of their office, and policy matters of long-term importance. It is also in this period that the assembly begins to set policy goals for their remaining time in office.

The duty to evaluate predecessors is an essential part of the retrospective accountability mechanism that incentivises assembly members to act for the long term. But it is also an essential part of the learning and socialisation process required by the Epistemological Criterion. The assemblies will learn much from the process of working to understand the successes, mistakes, and lessons learned by the previous generation. Moreover, the scoring rules that the assemblies use to evaluate the previous generation gives them a score card against which they can calibrate their own decisions, and see what choices of the previous generation were helpful or unhelpful to achieving a high score.

The score card that an assembly applies to the previous generation, and then uses to reflect on its own actions, is made up of a combination of high-level and low-level scoring rules. The high-level scoring rules are nigh-immutable goals that are sufficiently general to be applicable to every futures assembly, regardless of how the world changes. These scoring rules, then, are exactly the rules set out as duties in the previous section. Each assembly needs to ask of the previous generation: how well did the assembly fulfill its duty to exercise its powers and work with the government to produce policy that promotes the welfare of future generations? On the iterated variant, the assembly must also ask: how well did the assembly fulfill its duty to evaluate the decisions of one previous assembly and determine a reward for its members, as a function of their success in their duties? (This in turn will require examining the assembly two generations prior, to evaluate the accuracy of the previous assembly’s judgments.)

The low-level scoring rules are malleable and set by the assemblies themselves. These can contain any number of objectives, including: (1) high-level, abstract goals such as peace and security; (2) concrete, long-term metrics such as annual increase in GDP, target population levels, annual decrease in national Gini index, target levels of unemployment, target levels of atmospheric CO2 levels, or target levels of existential risk as reported in prediction markets; and (3) concrete, short-term goals and metrics such as international agreement to an OECD beneficial AI treaty, or the target size of the global nuclear arsenal. Each assembly inherits the low-level scoring rules from the assembly holding office just before it and has the flexibility to refine, change, or discard these rules in any way it pleases, as part of the learning phase.

The previous generation should be scored in part based on their success at achieving their own low-level scoring rules, but they should also be scored on their success at identifying the best low-level scoring rules for their assembly, creating an incentive for each assembly to not only act consistently with their scoring rules, but also to adopt ambitious and appropriate scoring rules for themselves.[[10]](#footnote-9)

#### 3. Structure

This section spells out one promising structure for the futures assembly. First is the machinery of the institution that helps it carry out its various duties: its archival arm, relationships with experts, research and investigation arm, and media and public relations team. Second is the description of the deliberating body of citizens itself: its constitution, size, salary, and term length.

#### 3.1. Archivists

The archivists are the assembly’s “institutional memory:” an independent bureaucratic body with no political status or access to the public, paid to (i) record all of the decisions and deliberations made by the assembly, (ii) package reports, evaluations, conference proceedings, and other materials for later assemblies, and (iii) chronicle the long-term social, economic, environmental, and political trends that are identified by the assembly as worthy of ongoing observation.

 The sole purpose of the archivist arm is to help solve epistemological obstacles faced by futures assemblies. By recording decisions and deliberations and packaging all of the ideas of the assembly in a way that makes them easy to absorb by later assemblies, they significantly reduce the epistemic burdens of retrospective accountability. By chronicling long-term trends, at the instruction of futures assemblies, they help solve “creeping problems”— the neglect of long-term issues, such as environmental issues, that happen as a consequence of an aggregate of many small, soritical changes over long time horizons (Glantz 1999).

#### 3.2. Experts

The experts serve to provide background seminars on a wide variety of issues relevant to assessing long-term priority areas and the long-term effects of policy. They also serve as consultants on the decisions and deliberations of the futures assembly. Experts should come from every discipline relevant to future-oriented policy, including the natural sciences, the social sciences, and philosophy, as well as relevant interdisciplinary endeavors such as futures studies, forecasting, and foresight. Their sole function is to infuse technocratic expertise into the deliberations of the assembly.

 Experts are generally academics, researchers, or public intellectuals with significant academic training. They are paid a stipend for their contributions and may be bought out from various teaching and research requirements, but they are not expected to support the futures assembly in a full-time capacity. Their term length can be the same as the term length of the futures assembly, or longer.

#### 3.3. Research and Investigation Arm

In addition to experts, the assembly also needs a research and investigation arm: a body of civil servants paid to carry out research directly commissioned by the futures assembly. This allows the assembly to expand its research and investigation capacity without requiring too much time from issue-area experts. The research and investigation arm should be as large as the futures assembly can productively make use of, potentially with many more members than the assembly itself. Each futures assembly may have substantial control over who is hired and fired in the research and investigation arm, but it should have an operations staff with little turnover, which can assist with the management and hiring of researchers to reduce the time costs of doing so to the futures assembly.

#### 3.4. Public Relations

It is important to ensure that the futures assembly is well-understood by and has strong communication with the general public, since institutional complexity is one of the most powerful determinants of public and elite refusal to invest in long-term policy priorities (Jacobs and Matthews 2012). Accordingly, all investigations, recommendations, meetings, and reports should be made available to the public and advertised widely by an internal public relations and media team. A primary goal of this project should be institutional transparency, with a secondary goal to craft compelling public narratives for the sake of movement-building and civic education. To ensure that the secondary goal does not become dominant, turning the PR team into a propaganda machine, among the public-facing documents should be the meeting notes taken by the independent archivists, which can be used to ensure the veracity of the PR team’s messaging.

#### 3.5. The Assembly

The most promising futures assemblies would be relatively large—c. 100–600 members, depending on the population of the jurisdiction in which it is implemented—to ensure demographic representativeness and resist corruption from interest groups. To further guard against corruption and ensure representativeness and minimal resignations, assembly members should be paid a high salary, for example commensurate with the typical salary for members of the national legislative body, plus the bonus already described, which should similarly be large enough to appropriately motivate assembly members.

Futures assembly members serve single terms and cannot be re-appointed, to avoid short-termist re-election incentives. The complexity of the job suggests a fairly long term length, to allow adequate time for learning and policymaking. On the other hand, the assembly members are not professional politicians and are selected from throughout the country, so the terms need to be minimally disruptive for randomly-chosen citizens who have other jobs and projects, suggesting a shorter term limit. Shorter term limits also restrict the extent to which assembly members can be reached by corrupting interest groups and allow more citizens to participate in the process over the long run. Nonetheless, a term length shorter than three years would not be adequate to the task of policymaking, so I suggest that a term length of 3–5 years is appropriate.[[11]](#footnote-10) The potential disruptiveness of this work should be ameliorated with opportunities to perform most work remotely and with generous family support policies.

#### 4. Checks and Balances

How will this new institution interact with the rest of the government? In light of its powers, tools, and policymaking role, the futures assembly will primarily engage with the legislature, regulatory executive agencies, and other public bodies, to precipitate policy changes at every level of government organisation. While the purpose of the futures assembly is to influence the rest of government, however, special attention should be given to ensure that the futures assembly is not unduly influenced by the present, short-termist government, especially the legislature. This is essential so that the futures assembly can insulate itself from the political business cycle and make reflective decisions about long-term matters outside of the news cycle and public attention. It is for this reason that the futures assembly is empowered to set its own research and policy agenda, and has its own research staff to investigate matters of long-term importance.

 The power of the futures assembly must of course be checked by the rest of the government so that it does not become excessively powerful. Fortunately this is no great risk, given that almost all of the powers proposed for the futures assembly herein have been advisory powers, and given that the futures assembly can at any time be repealed by the same legislative action that would give birth to it in the first place.[[12]](#footnote-11) The only substantial, binding power proposed for the futures assembly is the power to delay legislation (see ftn. 8), and this is to be checked either by the judiciary or by a supermajority of the legislature, as indicated previously.

#### 5. Conclusions

Political short-termism costs the global economy many billions and perhaps trillions of dollars annually and leads to many millions of deaths from disasters and suboptimal resource allocation. In this paper, I proposed that a futures assembly explicitly incentivised to promote the interests of future generations might be a promising strategy to ameliorate political short-termism. The most initially promising such retrospective accountability mechanism is an iterated mechanism, in which each generation of policymakers rewards the previous generation for their policy choices and for their own evaluation of the previous generation.

The representative mechanism proposed is among the most advanced and promising in the literature. But its potential for success is plagued with uncertainty. For one, while the mechanism is supported by contemporary political science and armchair theory, nothing like it has ever been implemented, and it is unclear whether the model’s assumptions about political motivation will hold up in practice. Secondly, I have not yet discussed the implementability of this reform. Given the widespread use of citizens assemblies and the creation of several future-focused institutions in the last two decades one should not assume that the proposal is nowhere feasible. But given that few *permanent* citizens assemblies exist, and that the accountability mechanism is novel, it is likely to be somewhat less implementable than other recent reforms.[[13]](#footnote-12)

Most importantly, while the mechanism may incentivise policymakers to act with the next hundred to one thousand years in view, it is very unlikely to be able to incentivise policymakers to act with the next million years in view. This is a problem, since very plausibly governments ought not be biased towards the near-term *at all*, and since vastly more people face the consequences of our actions over the next million years than will over the next thousand years, in expectation.

There is some reason to think that the interests of future generations over the next thousand years are highly correlated with the interests of future generations in a million years, since both are affected by the quality of intergenerational global public goods such as the environment, resources, technological progress, knowledge acquisition, and human survival—and in these domains, I think this much is true. But to the extent that the interests of near-future and far-future generations are decorrelated, it is very unlikely that the proposal succeeds in promoting the well-being of far-future generations. Future work could improve on the model proposed by extending the time-horizon of political decision-making even further, or by developing more narrowly-targeted mechanisms that incentivise policymakers to act specifically on policy objectives that benefit near- and far-future generations, such as permanently civilisation-ending catastrophes.

### Appendix: Proof of Conditions of Subgame Perfect Equilibrium

By Charlotte Siegmann

#### A1. Setup

There are infinitely many generations of decision-makers: i = 1, 2, 3, etc. Decision-maker i chooses $d\_{i }\in \{0,1\}$.$ d\_{i}=1$ is a policy that takes the interest of all future generations into account; $d\_{i}=0$ does not take their interest into account. For i, $ d\_{i}=1$ comes at a **cost C** due to selfish, political, and reputational preferences. For simplicity, I consider the binary choice. However, my conclusions do not change if d\_i is a continuum.

Once retired, decision-maker i may receive a reward of 0,$RD, RR,$ $RR +RD. $**RD is the reward for choosing** $d\_{i}=1. $RR is i’s reward if i rewarded i-1 adequately (**RR = the reward for adequate rewarding**). I assume constant marginal utility with respect to money.$ $

For i +1, the act of rewarding a sum of rewards $R\_{i}$ to decision-maker i comes at a personal benefit $B\_{i}= g\left(R\_{i}\right) with g'\left(R\_{i}\right)^{}^{}>1. R\_{ i} =RR\_{i} + RD\_{i}$. For simplicity, I assume the personal benefit is $g\left(R\_{i}\right) = γ⋅R\_{i}$ $with γ \in (0,1).$

The following behavioural strategy profile is called **iterated retrospective accountability (IRA)**: Everyone always plays $d\_{i}=1$. One gives the previous generation RD if and only if they play $d\_{i}$ $=1$. In addition, i rewards RR to i-1 if (i) $d\_{i-2}=1$ and i-1 has not rewarded RD to i -2 or (ii) $d\_{i-2}=0$ and i-1 has still rewarded them with RD - for more possible cases, see (John 2021).

The discount rate $δ$ includes pure time preferences (between decision-making and reward distribution ≈ 30 years), the probability of death before receiving one’s pension, and the probability the mechanism will get repealed. Conceivably, decision-maker s might not know whether past policies were myopic. $α$ is the probability that i+1 knows whether $d\_{i}=1$ or $d\_{i}=0 $happened. If$d\_{i}$ is unknown, i receives RD.

In the IRA equilibrium, decision-maker i has the following **utility function:**

$U\_{i}= -C ⋅ d\_{i}+ \left(d\_{i-1}\right) ⋅ γ⋅R\_{i-1} +\left(d\_{i}\right) ⋅ α ⋅ \left(1-δ\right) ⋅ RD\_{i}$ for i > 1

$ -C ⋅d\_{i}+ \left(d\_{i}\right) ⋅ α ⋅ \left(1-δ\right) ⋅ RD\_{i }$for i =1[[14]](#footnote-13)

I consider the conditions for IRA to be a subgame perfect Nash equilibrium.

#### A2. Reward Conditions for Subgame Perfection

First, $d\_{i}=1 is$ chosen if and only if C is smaller than the **net present value (NPV)** of the reward one receives for choosing $d\_{i}=1$.

$C \leq α ⋅ \left(1-δ\right) ⋅ RD$ (i)

$⇔RD \geq \frac{C}{α⋅\left(1-δ\right)}$ (ii)

I assume that when indifferent, decision-makers choose $d\_{i}=1. $(iii) shows the lowest RD such that $d\_{i}$ = 1 is induced.

$RD= \frac{C}{α⋅\left(1-δ\right)}$ . (iii)

However, because rewarding rewards make oneself better off, i+1 might give RD to i even if $d\_{i}=0$. To disincentivise this, the benefit of giving a reward must be smaller than the NPV of receiving RR, the reward for adequate rewarding, 30 years later:

*Utility Reward Distribution < NPV of RR*

$γ ⋅ RD \leq \left(1-δ\right) ⋅ RR$ (iv)

Combining (iii) with (iv), the reward for adequate rewarding must be at least:

$RR= \frac{γ⋅ C}{α⋅\left(1-δ\right)^{2}}$(v)

However, i+1 can only credibly threaten to withhold the RR from i if i+2 also has a credible threat, which requires:

$γ ⋅ RR \leq \left(1-δ\right)RD$ $⇔ γ \leq \left(1-δ\right)$(vi)

Equation (vi) is a key requirement. If (vi) is violated, the rewards will have to increase over time, making the mechanism and reward distribution unfeasible and uncredible.

For example, suppose decision-maker i (irrationally) chose $d\_{i}$=0 and gave RD to i-1 even though $d\_{i-1}=0$. In that case, decision-maker i gets a reward of zero. However, if she had not given a reward to i-1 ($d\_{i-1}=0)$ and chose $d\_{i}=1, $ then she receives RD + RR =$ \frac{C}{α⋅\left(1-δ\right)}+ \frac{γ⋅ C}{α⋅\left(1-δ\right)^{2}}$.

In sum, the following reward schedule sustains the equilibrium:

1. $ RD = \frac{C}{α⋅\left(1-δ\right)}$
2. $RR= \frac{γ^{ }⋅ C}{α⋅\left(1-δ\right)^{2}} $

#### A3. Minimal Rewards to Sustain Subgame Perfection

These rewards incentivise the correct non-myopic behaviour for rational agents. However, players are sometimes on off-equilibrium paths. Hence, sometimes no rewards will be given, even though they would be rational and warranted. This is a welfare loss. Hence, I compute the smallest possible rewards which sustain the desirable equilibrium.

Suppose individual x played $d\_{x}=0$. Then, as above, the lowest possible reward to make this decision irrational:

$RD\_{x}= \frac{C}{α⋅\left(1- δ\right)}$

If x+1 decides to give a reward to x anyway (because it makes x+1 better off), the smallest possible reward for x+1 for adequate rewarding, to make this action irrational is:

$RR\_{x+1}= \frac{γ^{ }⋅ C}{α⋅\left(1-δ\right)^{2}} $

However, suppose x+1 does reward x (despite $d\_{x}=0)$, and x+2 rewards x+1 for adequate rewarding. In this case, the minimum reward for adequate rewarding such that x+2 does not reward x+1:[[15]](#footnote-14)

$γ ⋅\frac{γ^{ }⋅ C}{α⋅\left(1-δ\right)^{2}}\leq \left(1-δ\right)⋅RR\_{x+2}$

$⇔ RR\_{x+2} \geq \frac{γ^{2}^{ }⋅ C}{α⋅\left(1-δ\right)^{3}}$

Suppose x plays $d\_{x}=0$. However, x+1 gave reward RD to x. In addition, x+2 to x+n reward RR to each previous generation. Hence, x+1, x+2, …, x+n all rewarded rewards incorrectly. However, if x+n is offered at least the following minimum reward for adequately rewarding RR, then the case is irrational for all players involved:

$RR\_{x+n}= \frac{γ^{n }⋅ C}{α⋅\left(1-δ\right)^{n+1}}$

### This minimum possible RR schedule also sustains the desirable equilibrium. As discussed above, if $γ>\left(1-δ\right),$ RR increases as n, the generational distance between the myopic decision-maker and the auditor increases. This makes the mechanism unsustainable. If $γ<\left(1-δ\right),$ RR decreases.

###

### Bibliography

Achen, C. H. and Bartels, L. M. (2002), ‘Blind retrospection: electoral responses to drought, flu, and shark attacks’ [Presentation], Annual meeting of the American Political Science Association, Boston.

Achen, C. H. and Bartels, L. M. (2016), *Democracy for realists: why elections do not produce responsive government* (Princeton University Press).

Aoki, R. and Vaithianathan, R. (2012), ‘Intergenerational voter preference survey – preliminary results’, CIS Discussion Paper Series.

Arnold, R. D. (1990), ‘*The logic of congressional action*’ (Yale University Press).

​​Banks, J. S. and Sundaram, R. K. (1993), ‘Adverse selection and moral hazard in a repeated elections model’, in W. A. Barnett, M. J. Hinich, and N. J. Schofield (eds.), *Political economy: institutions, competition, and representation* (Cambridge University Press), 295–311.

Barnovsky, A. (2006), ‘The ombudsman effect: how effective is the institution of the human rights commissioner in Russia’s regions?*’* Working paper. Available at: https://scholar.harvard.edu/files/alla-baranovsky/files/ombuds.paper\_.pdf (Accessed: 30th November, 2021).

Beck, N. (1982), ‘Does there exist a political business cycle: a Box-Tiao analysis’, in *Public Choice* 38: 205.

Beckman, L. and Uggla, F. (2016), ‘An ombudsman for future generations’, in I. González-Ricoy and A. Gosseries (eds.) *Institutions for Future Generations* (Oxford University Press).

Binder, S. A. (2006), ‘Can Congress legislate for the future?*’*, John Brademas Center for the Study of Congress, New York University, Research Brief 3.

Bohman, J. (2007), *Democracy Across Borders: From Dêmos to Dêmoi* (MIT Press).

Brennan, J. (2017), *Against Democracy* (Princeton University Press).

Brighouse, H. and Fleurbaey, M. (2010), ‘Proportionality and democracy’, in *Journal of Political Philosophy* 18: 137–155.

Caney, S. (2016), ‘Political institutions for the future: a five-fold package’, in I. González-Ricoy and A, Gosseries (eds.) *Institutions for future generations* (Oxford University Press).

Climate Assembly UK (2020), *The path to net zero: Climate Assembly UK full report*. Available at: https://climateassembly.uk/report/read/l/2.171.9mi5n.ywdwm.lqy35.2jmj7.html

Coleman, M., Devaney, L., Torney, D., and Brereton, P. (2019), *Ireland’s world-leading citizens’ climate assembly. What worked? What didn’t?* Climate Home News, June 27.

Elster, J. (2008), *Alchemies of the mind: rationality and the emotions* (Cambridge University Press).

Fearon, J. (1999), ‘Electoral accountability and the control of politicians: selecting good types versus sanctioning poor performance’, in A. Przeworski, S. C. Stokes, and B. Manin (eds.) *Democracy, accountability, and representation* (Cambridge University Press): 55–97.

​​Ferejohn, J. (1986), ‘Incumbent performance and electoral control’, in *Public Choice* 50:. 5–25.

Finkel, E. (2006), *Defending rights, promoting democracy*, The Institution of Ombudsman in Poland, Russia and Bulgaria, Working Paper.

Fiorina, M. P. (1977), ‘An outline for a model of party choice’, in *American Journal of Political Science* 21: 601.

Fishkin, J. S., Mayega, R. W., Atuyambe, L., Tumuhamye, N., Ssentongo, J., Siu, A. and Bazeyo, W. (2017), ‘Applying deliberative democracy in Africa: Uganda's first deliberative polls’, *Daedalus* 146:. 140.

Fishkin, J. S. and Luskin, R. C. (2005), ‘Experimenting with a democratic ideal: deliberative polling and public opinion’, *Acta Politica* 40: 284.

Gilligan, E. (2010), ‘The human rights ombudsman in Russia: the evolution of horizontal accountability’, in *Human Rights Quarterly* 32/3: 575.

Glantz, M. H. (1999), *Creeping environmental problems and sustainable development in the Aral Sea basin* (Cambridge University Press).

González-Ricoy, I. and Gosseries, A. (2016a), ‘Designing institutions for future generations’, in I. González-Ricoy and A. Gosseries (eds.) *Institutions for future generations* (Oxford University Press).

González-Ricoy, I. and Gosseries, A. (2016b), ‘An introduction’, in I. González-Ricoy and A. Gosseries (eds.) *Institutions for future generations* (Oxford University Press).

González-Ricoy, I. and Gosseries, A. (2016c), *Institutions for future generations* (Oxford University Press).

Greaves, H. and MacAskill, W. (2021), *The case for strong longtermism*. Available at: https://globalprioritiesinstitute.org/hilary-greaves-william-macaskill-the-case-for-strong-longtermism-2/ (Accessed: 30th November, 2021).

Grönlund, K., Herne, K., and Setälä, M. (2017), ‘Empathy in a citizen deliberation experiment’, in *Scandinavian Political Studies* 40: 457.

Healy, A. and Malhotra, N. (2009), ‘Myopic voters and natural disaster policy’, in *American Political Science Review* 103: 387.

Jacobs, A. M. and Matthews, J. S. (2012), ‘Why do citizens discount the future? Public opinion and the timing of policy consequences’, in *British Journal of Political Science* 42: 903.

John, T. M. and MacAskill, W. (2021), ‘Longtermist institutional reform,’ in T. M. John and N. Cargill (eds.) *The Long View* (London: FIRST).

John, T. (2020), ‘Longtermist institutional design and policy: a literature review,’ unpublished manuscript, URL: https://docs.google.com/document/d/1DdKjaRG0OOu7M1MB40bgdlJSykENQdoD8q28Yw9nuuc/

John, T. (2023), ‘Empowering Future Generations by Empowering the Young?’, in G. Bognar and A. Gosseries (eds.) *Ageing without Ageism? Conceptual Puzzles and Policy Proposals* (Oxford University Press).

Johnson, D. and Levin, S. (2009), ‘The tragedy of cognition: psychological biases and environmental inaction’, in *Current Science* 97: 1593.

Jones, N., O’Brien, M., and Ryan, T. (2018), ‘Representation of future generations in United Kingdom policy-making’, in *Futures* 102: 153.

Kamijo, Y., Teruyuki, T., and Yoichi, H. (2020), ‘Effect of proxy voting for children under the voting age on parental altruism towards future generations’, in *Futures* 122.

Key, V. O. Jr. (1966), *The responsible electorate: rationality in presidential voting 1936–1960* (Harvard University Press).

Koo, J. and Ramirez, F. O. (2009), ‘National incorporation of global human rights: worldwide expansion of national human rights institutions, 1966-2004’, in *Social Forces* 87/3: 1321.

List, C., Luskin, R. C., Fishkin, J. S. and McLean, I. (2013), ‘Deliberation, single-peakedness, and the possibility of meaningful democracy: evidence from deliberative polls’, in *The Journal of Politics* 75: 80.

MacKenzie, M. K. (2016), ‘Institutional design and sources of short-termism’, in I. González-Ricoy and A. Gosseries (eds.) *Institutions for Future Generations* (Oxford University Press).

Norwood, F. B., Tonsor, G. and Lusk, J. L. (2019), ‘I will give you my vote but not my money: Preferences for public versus private action in addressing social issues’, in *Applied Economic Perspectives and Policy* 41: 96.

Ord, T. (2020), *The precipice: existential risk and the future of humanity* (Hachette Books).

Paul, A. S., Lusk, J. L., Norwood, F. B. and Tonsor, G. T. (2019), ‘An experiment on the vote-buy gap with application to cage-free eggs’, in *Journal of Behavioral and Experimental Economics* 79: 102.

Reif, L. C. (2011), ‘Transplantation and adaptation: the evolution of the human rights ombudsman’, in *Boston College Third World Law Journal* 31: 269.

Rogoff, K. (1990), ‘Equilibrium political budget cycles’, in *American Economic Review*, 80: 21–36.

Spitzer, R. J. (1988), *The presidential veto* (SUNY Press).

Steiner, J., Bächtiger, A., Spörndli, M. and Steenbergen, M. R. (2005), *Deliberative politics in action: analyzing parliamentary discourse* (Cambridge University Press).

Tufte, E. R. (1978), *Political Control of the Economy* (Princeton University Press).

Weber, E. U. (2006), ‘Experience-based and description-based perceptions of long-term risk: why global warming does not scare us (yet)’, in *Climatic Change* 77: 103.

1. This paper was greatly improved by feedback from Greg Bognar, Mark Budolfson, Adam Gibbons, Axel Gosseries, Alex Guerrero, Charlotte Siegmann, Larry Temkin, and one anonymous referee. [↑](#footnote-ref-0)
2. For a contrary view, see (Beck 1982). [↑](#footnote-ref-1)
3. Retrospective accountability has a similar incentives structure to electoral democracy. As some contemporary theorists of democratic accountability claim, democracies make policymakers accountable to the public through a process of “retrospective voting” (Key 1966, Fiorina 1981, Ferejohn 1986, Rogoff 1990, Banks and Sundaram 1993, Fearon 1999). However, unlike electoral democracy, retrospective democracy has much longer time horizons, and offers a much tighter connection between decisions and rewards (assembly members are audited *after* they leave office). [↑](#footnote-ref-2)
4. Note that this condition is substantially weaker than common knowledge. [↑](#footnote-ref-3)
5. The precise level of confidence required can be derived from the auditor’s own preferences about whether to cooperate or defect and the size of the reward for cooperation. If the auditor is intrinsically motivated to cooperate, the auditor does not need to be very confident that the subsequent auditor will cooperate. If the auditor is intrinsically motivated to defect, they need to be more confident that the subsequent auditor will cooperate, or the size of the reward for cooperating needs to be larger. In general, as long as each auditor believes that there is a better-than-chance likelihood of being rewarded for cooperation, cooperation can be sustained for an arbitrarily large reward. [↑](#footnote-ref-4)
6. An anonymous reviewer of this chapter noted that some political scientists (such as Achen and Bartels 2002, 2016) are skeptical of retrospective voting as it normally functions in democracy, in that they believe that voters are too ignorant to accurately apportion credit to previous administrations and are therefore swayed in their voting by entirely irrelevant, emotionally salient information (e.g. recent shark attacks). However, the most important “finding” in this literature has failed to replicate (Fowler and Hall 2018), and the epistemic situation of members in the proposed assembly is in many ways superior to the epistemic situation of members of an electoral democracy. After all, retrospective voting is these members *entire jobs* for multiple years (2.2), the members have much better access to information due to support from experts and archivists (3.1–3.3), and they are less polarised than ordinary members of the demos due to effects from sortition and deliberation (1.1). [↑](#footnote-ref-5)
7. In particular, any promising futures assembly must be empowered to review the scoring rules that are used to determine the bonuses of the previous assemblies, along with the details of the rules about how to archive information and select experts. As each of these sets of rules have a powerful effect on the incentives of the institution, I suggest that the power of review can only be used in such cases if two successive assemblies achieve supermajority support for the proposed rule changes, and that the changes to the assembly’s rules only go into effect when a third successive assembly comes into office. This system of review reduces groupthink, prevents assemblies from changing the rules to their own advantage, and ensures the assemblies are not entirely self-governing. [↑](#footnote-ref-6)
8. That said, there are several reasons we might consider expanded powers for a futures assembly: it may be more feasible in the future to sustain a robust, future-oriented political institution, and the assembly may need expanded powers to fully achieve its policy goals or for bargaining or self-protection. What sort of “harder powers'' might futures assemblies ideally possess? Assemblies might, as a few examples, (1) require the legislature to read and respond to the assembly’s reports and recommendations; (2) require the government to notify the public about a proposed policy; (3) require “posterity impact assessments'', or reports on the future social or economic impact of a policy proposal (John and MacAskill 2021); (4) have the power to directly place a piece of proposed legislation in front of a legislative body and force a vote; (5) have the power to delay legislators from passing a potentially harmful piece of legislation (reviewable by judiciary or supermajority of legislatures, to prevent the assembly from obstructing a genuinely urgent policy need). With these quasi-legislative powers, the assembly might aim to precommit governmental bodies into taking future actions which will later seem difficult but high impact, to extend or eliminate budget windows, or to force policymakers to amend their implicit discount rate on utility (for an overview of why non-zero discount rates on utility are ethically unjustifiable, see Greaves 2017). [↑](#footnote-ref-7)
9. It might be too epistemically demanding for any small group of citizens to make informed, sensible decisions across all domains. One tentative solution is to divide the intellectual labor *temporally*—e.g., giving the first assembly a generalist focus, which can in turn determine a particularly pressing policy priority for the second assembly to specialise in, then rotating back to a generalist focus for the third assembly, and so on. [↑](#footnote-ref-8)
10. One crucial question is whether assemblies should evaluate previous generations from an *ex post* perspective — how good were the assembly’s decisions, given how the world turned out?—or an *ex ante* perspective—how good were the assembly’s decisions, given the information available to them at the time? An *ex ante* system seems more epistemically tractable for the previous generation, whereas an *ex post* system seems more epistemically tractable to the evaluating generation. Neither seems unqualifiedly preferable to the other. I suggest that whichever system is adopted properly balances the epistemic tractability burdens between both generations, and properly incentivises the prior generation to seek out new information. [↑](#footnote-ref-9)
11. An anonymous referee suggests that the need for selected individuals to learn everything relevant to their job counts in favour of asynchronous terms, so that at any given time there are new members and more experienced members who can teach the new members. This is a plausible suggestion, but it also makes it more likely that the futures assembly develops a distinctive internal culture that is transmitted from cohort to cohort, which could make its culture deviate increasingly from the surrounding political culture. This may make the assembly’s epistemic and motivational features more locked in and less susceptible to change with novel cohorts. It is not clear these features would affect the success of the scheme overall. [↑](#footnote-ref-10)
12. Given the possibility of repeal, the relevant jurisdiction should adopt measures to ensure that the futures assembly will still be evaluated and paid by an ad hoc body of auditors at the appropriate time, even though the institution has itself been repealed. This will increase every assembly’s confidence in the persistence of the incentives mechanism, increasing their willingness to pay personal costs for the sake of future generations, given the higher guarantee of a reward for doing so. [↑](#footnote-ref-11)
13. Some political theorists have discussed this issue under the label “the bootstrapping problem.” As the argument sometimes goes, there is something odd or surprising about *any* proposed reform that asks citizens to vote for a policy that then binds them to take more action than they would otherwise be willing to do. But these political theorists overestimate the challenge. It is a widespread feature of tax policy (e.g. redistributive policy) and consumer policy (e.g. cage-free egg policy) that citizens who vote for these policies do so even though they would not voluntarily increase their spending on government programs or voluntarily reduce their consumption of goods with negative externalities. However puzzling this is from the armchair, it is a well-established empirical fact. [↑](#footnote-ref-12)
14. No reward for adequate rewarding because there was no previous policymaker. [↑](#footnote-ref-13)
15. If < 1-δ, this is lower than the reward in the original model. [↑](#footnote-ref-14)