Existential Risk and Equal Political Liberty

Abstract: Rawls famously argues that the parties in the original position should agree upon the two principles of justice, thus guaranteeing citizens equal political liberty. We argue on the contrary that the parties have reason to reject the requirement of equal political liberty. By Rawls' own lights, the parties must be greatly concerned to mitigate existential risk. But it is doubtful whether democracies optimally do so. Indeed, no one currently knows which political systems would. Consequently, the parties have reason to reject the requirement of equal political liberty in favor of an experimentalist political approach which does not rule out non-democratic systems which might mitigate existential risk optimally. We begin by summarizing some general facts about existential risk and three pathologies of democracy which hinder democracy's risk-mitigating capacities: voter ignorance, voter irrationality, and short-termism. We then argue that these facts, along with the possibility of less pathologized alternatives to democracy, give the parties reason to reject the requirement of equal political liberty. After addressing some further aspects of Rawls' theory, we consider two objections to our claim that it is doubtful whether democracies optimally mitigate existential risk. We conclude with a brief discussion of our argument's broader implications.

Keywords: democracy, existential risk, long-termism, political experimentalism, Rawls

Existential Risk and Equal Political Liberty

Nowhere in his substantial body of work does Rawls address existential risk—risk of catastrophic events which would permanently destroy humanity's long-term potential. In fact, scarcely anything has been written on existential risk in the vast secondary literature on Rawls and his theory of justice. This silence is unfortunate. It is hard to deny that the parties in the original position would be greatly concerned to mitigate existential risk, which threatens the lives and fundamental interests of countless generations. Indeed, the parties represent "continuing persons" spanning multiple generations who must agree to principles of justice which they would want all previous generations to have followed.[1] Presumably, no generation would want previous generations to have followed principles of justice which, by failing to mitigate (or even exacerbating) existential risk, threatened its very existence. But then one of the parties' greatest concerns would be to agree upon principles of justice which did not hinder our capacity to mitigate existential risk.

What, however, do principles of justice have to do with existential risk? Rawls famously argues that the parties would agree upon the two principles of justice. Among other things, these principles guarantee citizens "the same indefeasible claim to a fully adequate scheme of equal basic liberties."[2] This scheme includes equal political liberty (hereafter "EPL"), which grants citizens an equal right to vote—specifically, "one person one vote"—and "equal access … to

[1] John Rawls*, A Theory of Justice: Revised Edition* (Cambridge, MA: Harvard University Press, 1999), 118; John Rawls*, Justice as Fairness: A Restatement* (Cambridge, MA: Harvard University Press, 2001), 160.

[2] Rawls*, Justice as Fairness*, 42.

public office."[3] But would political systems which grant citizens EPL—in other words, democracies—optimally mitigate existential risk?[4] In this paper, we argue that it is doubtful whether they would. Indeed, no one currently knows which political systems would do so. The parties in the original position therefore have reason to reject EPL as a requirement of justice so as not to rule out non-democratic political systems which might optimally mitigate existential risk.

We begin in Section I with a survey of some of the relevant literature on existential risk. We then discuss a substantial body of work which shows that voters and other democratic political actors are prone to ignorant, irrational, and short-termist decision-making. These pathologies, we claim, hinder democracies' capacities to deal with complex problems like existential risk. Lastly, we close by outlining an alternative approach to political justice which we call *political experimentalism*. This approach to political justice permits political experimentation to determine which political systems best promote our various ends. Importantly, political experimentalism allows for experimentation with some non-democratic political systems which might optimally mitigate existential risk.

In Section II, we examine the parties' deliberations in light of the relevant general facts about existential risk and the pathologies of democracy. As described by Rawls, the parties "know whatever general facts affect the choice of the principles of justice."[5] Hence they know that some non-democratic systems might mitigate existential risk better than democracies. They

[3] Rawls, *Theory*, 196, 203.

[4] Following Rawls, we use "democracy" to refer to any political system which grant citizens EPL—in other words, any system with equal and universal suffrage.

[5] Ibid., 119.

therefore have reason to reject EPL as a requirement of justice in favor of political experimentalism.

In Section III, we address two objections to our claim that it is doubtful whether democracies optimally mitigate existential risk. The first—the objection from epistemic democracy—is that democracies actually outperform other political systems by drawing upon the collective intelligence of crowds. The second—the objection from democratic reform—is that suitably reformed democracies *would* optimally mitigate existential risk even if actual democracies do not. In reply, we argue that both of these objections underestimate the extent to which ignorance, irrationality, and short-termism pathologize decision-making in both actual and hypothetical reformed democracies.

We conclude in Section IV with a brief discussion of our argument's broader implications. Among other things, our argument can be generalized beyond Rawls himself to anyone with a similar commitment to mitigating existential risk. Since most of us do in fact share such a commitment, most of us have at least some reason to reject EPL as a requirement of justice.

Two caveats before we proceed. First, we do not claim to know which political systems would in fact optimally mitigate existential risk. Indeed, we do not claim to *know* that democracies would not do so. We argue only that it is doubtful whether they always would.[6] Second, we do not claim to know whether equal political liberty should be rejected as a

[6] More precisely, we argue that it is doubtful whether they always would *at least* under the circumstances in which the two principles are paradigmatically operative—roughly, circumstances of relative affluence. Rawls endorses S.I. Benn's suggestion that the two principles "come into their own" only under "conditions of affluence" in which "everyone's generally agreed needs" are provided for (*Theory*, xx; S.I. Benn, "Egalitarianism and the Equal Consideration of Interests," in *Justice and Equality*, ed. Hugo Bedau (Englewood Cliffs, NJ: Prentice Hall, 1971), 166-67). Under less favorable conditions, a more general conception of justice may be operative instead. We discuss the relation between this more general conception and the two principles further in Section II.

requirement of justice *all things considered*. It is beyond the scope of this essay to evaluate all the considerations which have been proposed in favor of accepting that requirement. We argue only that the parties' interest in mitigating existential risk gives them significant reason to reject it.

I. Existential Risk, the Pathologies of Democracy, and Political Experimentalism

I.I. Existential Risk

What is existential risk (hereafter "x-risk")? Following Toby Ord, we say that x-risks are risks of existential catastrophes which would destroy humanity's long-term potential.[7] Ord captures this long-term potential well:

> Humanity is about two hundred thousand years old. But the Earth will remain habitable for hundreds of millions more—enough time for millions of future generations … enough to create heights of flourishing unimaginable today. And if we could learn to reach out further into the cosmos, we could have more time yet: trillions of years, to explore billions of worlds. Such a lifespan places present-day humanity in its earliest infancy. A vast and extraordinary adulthood awaits.[8]

X-risk, however, threatens to rob us not just of an extraordinary adulthood but of *any* future whatsoever. For anyone who wishes to safeguard the future of humanity (and of terrestrial life itself), the need to mitigate x-risk is clear.

[7] Toby Ord, *The Precipice: Existential Risk and the Future of Humanity* (New York: Hachette Books, 2020). For an overview of different definitions of x-risk in the relevant literature, see Phil Torres, "Existential Risks: A Philosophical Analysis," *Inquiry* (2019), https://doi.org/10.1080/0020174X.2019.1658626. Note that our argument does not depend on which definition is used.

[8] Ibid., 3.

Paradigmatic examples of possible existential catastrophes include natural extinction events (such as asteroid impacts, naturally occurring pandemics, supervolcanic eruptions, and stellar explosions), anthropogenic extinction events (such as nuclear holocausts and engineered pandemics), and extinction events arising from both natural and anthropogenic factors (such as tail risks associated with runaway greenhouse effects).[9] But existential catastrophes need not involve extinction. On Ord's account, the emergence of a dystopia from which we could not recover would also constitute an existential catastrophe, even if it did not literally wipe us out.[10]

Naturally, humanity is interested in mitigating risks of all sorts. But x-risks are especially grave. Civilization can recover from non-existential catastrophes. A true existential catastrophe, however, precludes any possibility of recovery. Mitigating x-risk is therefore a wholly proactive endeavor.

In assessing different political systems' capacities to mitigate x-risk, we should bear in mind three features of most x-risks:

- *Long timescales*: Most existential catastrophes are unlikely to occur for many thousands—or even millions—of years. For example, an asteroid impact may not threaten humanity with extinction for several million years, because there is an inverse relationship between asteroid size and frequency of impact (with more dangerous impacts occurring much less frequently).

[9] See Nick Bostrom and Milan M. Ćirković, "Introduction," in *Global Catastrophic Risks*, eds. Bostrom and Ćirković (Oxford: Oxford University Press, 2008); Phil Torres, *Morality, Foresight, and Human Flourishing: An Introduction to Existential Risks* (Durham, NC: Pitchstone Publishing, 2017); Ord, *Precipice*.

[10] Ord, *Precipice*, 145-55. For relevant discussion, see Bryan Caplan, "The Totalitarian Threat," in *Global Catastrophic Risks*, eds. Bostrom and Ćirković.

- *Low probabilities*: Relatedly, most existential catastrophes are individually unlikely. For instance, the probability of an existentially catastrophic stellar explosion within the next century is only 1 in 1,000,000,000.[11]
- *Complexity*: Most x-risks cannot be adequately understood without a good grasp of complex technical subjects. For example, the risk of value-misaligned artificial intelligence cannot be adequately understood—let alone mitigated—without a good grasp of computer science, decision theory, and other cognitively demanding fields of study.

In short, most x-risks involve far-off, low-probability, and complex events. Although the individual probability of any one such event within the next several decades may be quite low, the *cumulative* probability—the total x-risk—may be worryingly high.[12] X-risk is therefore both especially hard and especially important to mitigate.

It is impossible to study directly which political systems optimally mitigate x-risk. For obvious reasons, we cannot wait until *after* an existential catastrophe has occurred to learn from experience which political systems deal with it best. Consequently, in assessing different systems' relative capacities to mitigate x-risk, the best we can do is to study the extent to which they promote informed, rational, and long-term decision-making *in general*. Plausibly, systems which do not effectively promote such decision-making are ill-suited to deal with problems like x-risk. Of course, such an indirect method can hardly be definitive—one reason we do not claim to *know* which systems would optimally mitigate x-risk. But (as we hope to show) it can still be quite fruitful.

[11] Ord, *Precipice*, 167.
[12] Ibid.

I.II. Three Pathologies of Democracy

Are democracies best suited to deal with problems like x-risk involving far-off, low-probability, or complex events? We very much doubt so, because democratic decision-making is compromised by at least three pathologies: voter ignorance, voter irrationality, and democratic short-termism.[13]

First, democratic decision-making is compromised by *voter ignorance*. Since becoming politically well-informed is highly costly and only minimally beneficial to individual voters, most democratic voters are rationally ignorant.[14] Decades of research confirm that typical voters are ignorant even of basic facts about the structure and function of political institutions, the identity and platforms of political candidates, and much more.[15] Unsurprisingly, most voters are also ignorant of important social-scientific subjects relevant to democratic politics—not to mention the many complex subjects relevant to x-risk mitigation.

Voters' widespread ignorance has two mutually reinforcing consequences: First, ignorant voters often support candidates endorsing harmful policies; second, both prospective and current

[13] Even accounting for these three pathologies, it might still be wondered whether democracy is *less* pathologized than other political systems overall, because current democracies arguably counteract *other* pathologies (like corruption) more effectively than current non-democracies. For reasons which we discuss further in Section III.I, we doubt whether current democracies' success in counteracting such pathologies is best explained by their democratic institutions *themselves*. But even if it is, the possibility remains that some non-democratic systems would counteract these pathologies *even more* effectively. Since we cannot discount this possibility, and since democracy exhibits unique pathologies of its own, we doubt whether democracies are *best* suited to deal with problems like x-risk.

[14] Anthony Downs, *An Economic Theory of Democracy* (New York: Harper and Row, 1957). But see also Paul Gunn, "Looking But Not Seeing: The (Ir)relevance of Incentives to Political Ignorance," *Critical Review: A Journal of Politics and Society* 27, nos. 3-4 (2015): 270-90; Jeffrey Friedman, *Power Without Knowledge: A Critique of Technocracy* (Oxford: Oxford University Press, 2019).

[15] For overviews of the literature on voter ignorance, see Bryan Caplan, *The Myth of the Rational Voter: Why Democracies Choose Bad Policies* (Princeton, NJ: Princeton University Press, 2007); Ilya Somin, *Democracy and Political Ignorance: Why Smaller Government Is Better* (Stanford, CA: Stanford University Press); Jason Brennan, *Against Democracy* (Princeton, NJ: Princeton University Press, 2016).

legislators are incentivized to respond to ignorant voters' preferences.[16] The joint effect of these two consequences is the frequent implementation of laws and policies which go against citizens' interests—including their interest in mitigating x-risk. A salient recent example is many democracies' ineffective response to the COVID-19 pandemic.[17] Notably, if COVID-19 had been much deadlier, the ensuing pandemic could have become a genuine existential catastrophe for which most democracies would have been terribly underprepared.

Second, and similarly, democratic decision-making is compromised by *voter irrationality*. Just as becoming politically well-informed is highly costly and only minimally beneficial to individual voters, so too is conforming to normal standards of epistemic rationality in political belief formation. In fact, in many partisan environments, epistemic rationality can even be *penalized*. For instance, within some ingroups, rationally moderating one's beliefs may result in ostracization and other social costs. Hence most democratic voters behave in paradigmatically epistemically irrational ways in the political domain. Indeed, most voters are *rationally irrational*: (practically) rational in their (epistemic) irrationality.[18] Naturally, rationally irrational voters incentivized to form irrational political beliefs are not especially well suited to deal with political problems of any kind. X-risk is no exception.

Third, and maybe most importantly, democratic decision-making is compromised by *short-termism*. A large body of work in political science suggests that democracies focus unduly

[16] Of course, legislators are not *fully* responsive to voters' preferences. But voters still exert *some* influence over laws and policies. For further discussion, see Christopher H. Achen and Larry M. Bartels, *Democracy for Realists: Why Elections Do Not Produce Responsive Government* (Princeton, NJ: Princeton University Press, 2016), 318-19.

[17] See, e.g., Eric Winsberg, Jason Brennan, and Chris W. Surprenant, "How Government Leaders Violated Their Epistemic Duties During the SARS-CoV-2 Crisis," *Kennedy Institute of Ethics Journal* 30. no. 3 (2020): 215-42.

[18] Caplan, *Myth of the Rational Voter*.

on short-term problems at the expense of long-term ones.[19] Of course, short-termism is not a problem only for democracies. Some determinants of short-termism are general and pose a challenge for all political systems. For example, many cognitive biases can lead us to neglect long-term issues. In conditions of informational uncertainty about the future, we often discount the value of actions with long-term benefits relative to actions with more certain short-term benefits.[20] In addition, we are often more responsive to salient and visible risks than to risks apparent only from abstract reflection or extrapolation from data.[21] But salient, visible, and short-term risks are not necessarily the most threatening ones, and in any case most x-risks are neither salient, visible, nor short-term. Thus most members of *any* political system can be expected to neglect long-term problems like x-risk, because psychological determinants of short-termism predispose them to biased short-term thinking.

More striking than such psychological determinants of short-termism, however, are the *institutional* determinants of short-termism specifically in democracies.[22] These determinants prevent the formation and implementation of long-termist policy, undercut political actors' motivation to mitigate long-term risk, and hinder our capacity to gather information about such risks and reason appropriately about them. If democratic institutions *themselves* further

[19] For an overview of the relevant literature, see Tyler M. John and William MacAskill, "Longtermist Institutional Reform," in *The Long View*, eds. John and Cargill (London: FIRST, 2021). See also Simon Carey, "Political Institutions for the Future: A Five-fold Package" and Iñigo González-Ricoy and Axel Gosseries, "Designing Institutions for Future Generations: An Introduction," in *Institutions for Future Generations*, eds. González-Ricoy and Gosseries (Oxford: Oxford University Press, 2016); Alan M. Jacobs, "Policy Making for the Long Term in Advanced Democracies," *Annual Review of Political Science* 19 (2016): 433-54; Graham Smith, *Can Democracy Safeguard the Future?* (Medford, MA: Polity Press, 2021).

[20] Shane Frederick, George Loewenstein, and Ted O'Donoghue, "Time Discounting and Time Preference: A Critical Review," *Journal of Economic Literature* 40, no. 2 (2002): 351-401; Yoram Halevy, "Strotz Meets Allais: Diminishing Impatience and the Certainty Effect," *American Economic Review* 98, no. 3 (2008): 1145-62.

[21] Elke U. Weber, "Experience-based and Description-based Perceptions of Long-Term Risk: Why Global Warming Does Not Scare Us (Yet)," *Climatic Change* 77 (2006): 103-20. Relatedly, people's responsiveness to risks is often disproportionate to the threat they pose, especially when they involve large numbers or hard-to-calculate probabilities.

[22] John and MacAskill, "Longtermist Institutional Reform."

incentivize us to neglect the long term, then democracies will arguably mitigate x-risk less effectively than other (more long-termist) political systems.

Foremost among institutional determinants of short-termism in democracies are electoral incentives. Because politicians want to be (re-)elected, they tend to prioritize policies which offer constituents visible short-term benefits, since they can benefit politically from implementing such policies while imposing their costs on later generations who cannot sanction them.[23] But electoral incentives are far from the only institutional determinants of short-termism in democracies. Politicians also have *financial* incentives to be short-termist, because they are often economically dependent on organizations which can influence them to focus on the short term.[24] Additionally, policymakers often rely upon performance indicators which further incentivize short-termism by tracking performance over short timespans, using short budget windows, and so on.[25] These and other institutional determinants strongly incentivize democratic political actors to neglect long-term problems—including x-risk.

If democracies are pathologized by voter ignorance, voter irrationality, and short-termism, then we should expect few (if any) democracies to prioritize the goal of effective x-risk mitigation. Even if some do, we should expect the complexity of that task to keep most voters from forming the appropriate policy preferences. Ultimately, voters do not look very far

[23] W.D. Nordhaus, "The Political Business Cycle," *Review of Economic Studies* 42, no. 2 (1975): 169-90; Sarah Binder, "Can Congress Legislate for the Future?", in *John Brademas Center for the Study of Congress: Research Brief 3* (2006).

[24] Carey, "Political Institutions for the Future."

[25] Binder, "Can Congress Legislate for the Future?"; John and MacAskill, "Longtermist Institutional Reform."

ahead (or behind) at the ballot box.[26] We can hardly expect them to worry about stellar explosions a million years hence—let alone the next pandemic, which could come at any time.[27]

I.III. Political Experimentalism

Of course, democracy's pathologies do not *entail* the suboptimality of democratic x-risk mitigation. No one can definitively show that a political system is optimal or not in some respect without exhaustive comparative institutional analysis—which is currently impossible. Nonetheless, the pathologies of democracy undeniably give the parties in the original position *some* reason to doubt democracies' risk-mitigating capacities and to consider a more experimentalist approach to political justice. Why this is so is perhaps best illustrated with an analogy.

Suppose that Forrest is interested in betting on some upcoming horse races of different distances and kinds: ten-furlong races, harness races, and so forth. Forrest has only minutes to place his bets before the books close, so he cannot research the entire field in the upcoming races beforehand. As Forrest is deliberating on his bets, he sees a palomino colt gingerly hobble by. Three of its legs are in casts.

Forrest has only limited and general information about which horses to bet on, some of which suggests that no *one* horse will necessarily be favored to win all the upcoming races. (Among other things, different breeds of horses are best suited to different kinds of races.)

[26] Achen and Bartels, *Democracy for Realists*.

[27] Some x-risks (such as risks of stellar explosions) cannot currently be mitigated. Hence it might be argued that democracy's failure to mitigate these risks does not count against democracy itself, since any other political systems would fail similarly. However, even if *some* x-risks cannot currently be mitigated, others still can be, and it is doubtful whether democracies optimally mitigate even these other risks. Moreover, a robust x-risk mitigation strategy plausibly includes efforts to develop *new capacities* to mitigate currently unmitigable x-risks, and most democracies fail to do even that.

Forrest therefore has at least some reason not to bet on just one horse in the upcoming races—especially the palomino colt, because of its visible injuries. To be sure, it is *possible* that all the other horses in the field are in even worse condition than the colt. Nevertheless, unless Forrest has some particular reason to think that *all* the other horses are *that* systematically injury-prone, it is reasonable for him to doubt whether the palomino colt is favored to win all the upcoming races.

The parties' position is not unlike Forrest's. Just as Forrest cannot research and observe all the horses in the field, so too the parties (and we) cannot directly research and observe all the innumerable possible political systems. Like Forrest, the parties have only limited and general information about which political systems to "bet" on, and some of their information suggests that no one political system will *always* mitigate x-risk optimally. A longstanding view in the social sciences is that the laws and political institutions best suited to promote a society's ends *vary* with its "climate, geological features, economic characteristics, religion, customs, etc."[28] The parties therefore have at least some reason to doubt whether any one political system would always optimally mitigate x-risk—including democracy, because of its documented pathologies. To be sure, it is *possible* that all other political systems would mitigate x-risk even worse (or no better) than democracies do. Nevertheless, unless the parties have some particular reason to think that *all* possible political systems are *that* systematically pathologized, it is reasonable for them to doubt whether democracies always mitigate x-risk optimally.

[28] George Klosko, "Rawls' Argument from Political Stability," *Columbia Law Review* 94, no. 6 (1994): 1882-1897, 1891. See also Alex Guerrero, "Political Functionalism and the Importance of Social Facts," in *Political Utopias*, eds. Vallier and Weber (Oxford: Oxford University Press, 2017); Guerrero 2017; Paul Dragos Aligica, Peter J. Boettke, and Vlad Tarko, *Public Governance and the Classical-Liberal Perspective: Political Economy Foundations* (Oxford: Oxford University Press, 2019); etc. As we point out in Section II.I, Rawls himself endorses such a view of economic systems (*Theory*, 242).

If democracies do not optimally mitigate x-risk, which political systems do? We do not claim to know (nor does our argument require us to).[29] In fact, as should now be apparent, we doubt whether any *one* political system would always do so. Furthermore, even if one system would, the only way to identify it would be to *experiment* with different political systems and compare their capacities to deal with long-term, complex problems. Accordingly, as we argue further in Section II.I, the parties have reason to embrace what we call *political experimentalism* and to rule in multiple political systems as possibly just.

Political experimentalism is an approach to political justice which permits experimentation with a range of different political systems rather than ruling out almost all possible systems in advance. Thus far, human societies have experimented with only a few political systems, and it is doubtful whether any of these systems is optimal in every respect. However, the only definitive way to determine the advantages of *other* political systems is to experiment with them. Consequently, experimentalism permits (and even encourages) experimentation with different political systems to determine which ones are best suited to promote our political ends under different circumstances.

It is beyond the scope of this essay to offer a systematic account and defense of political experimentalism. Nonetheless, we believe that the problem of x-risk clearly illustrates the advantages of an experimentalist approach. Because the parties cannot determine in advance which political systems optimally mitigate x-risk, they have at least some reason to rule in multiple political systems as possibly just. Doing so allows different societies to determine

[29] Rawls does not require his own argument for the two principles to "constructively characterize or enumerate all possible conceptions of justice" (*Theory*, 107). So he can hardly require our argument to constructively characterize or enumerate all possible political systems.

through experimentation which political system deals best under their various circumstances with problems like x-risk.

Naturally, until we have experimented more widely with different political systems, we can only speculate as to their relative capacities to deal with such problems. Nevertheless, it seems likely to us that some non-democratic systems *would* in fact mitigate x-risk better than democracies, either by incentivizing a greater focus on the long term or by reducing the harm of voter ignorance and irrationality (or both). Perhaps some *epistocratic* systems would mitigate x-risk better than democracies under certain circumstances by reducing the political influence of ignorant and irrational voters, or by changing political selection mechanisms to promote the selection of more competent political leaders.[30] Under other circumstances, some *lottocratic* systems might excel by removing harmful short-termist electoral incentives.[31] It may be that a *hybrid* political system combining features of different forms of government would optimally mitigate x-risk, or that some new and hitherto unconceived system would fare best. Perhaps, as some argue, only a sufficiently advanced artificial intelligence could effectively steward humanity into the far future.[32] Again, we do not claim to know which political systems would

[30] Daniel A. Bell, *The China Model: Political Meritocracy and the Limits of Democracy* (Princeton, NJ: Princeton University Press, 2015); Tongdong Bai, *Against Political Equality: The Confucian Case* (Princeton, NJ: Princeton University Press, 2020). For a list of several possible epistocratic systems, see Brennan, *Against Democracy*. For further discussion, see Thomas Mulligan, "Plural Voting for the Twenty-First Century," *The Philosophical Quarterly* 68, no. 271 (2018): 286-306; Adam Gibbons, "is Epistocracy Irrational?", *Journal of Ethics and Social Philosophy* (forthcoming): 19-24. Note that we elide here the distinction sometimes drawn between epistocratic and meritocratic political systems.

[31] Alex Guerrero, "Against Elections: The Lottocratic Alternative," *Philosophy and Public Affairs* 42 (2014): 135-78.

[32] For one example of a hybrid political system, see Nicolas Berggruen and Nathan Gardels, *Intelligent Governance for the 21st Century: A Middle Way Between West and East* (Cambridge, UK: Polity, 2013). For discussion of the possible role of artificial intelligence in x-risk mitigation, see Nick Bostrom, "What Is a Singleton?", *Linguistic and Philosophical Investigations* 5, no. 2 (2005): 48-54; Phil Torres, "Superintelligence and the Future of Governance: Prioritizing the Control Problem at the End of History," in *Artificial Intelligence Safety and Security* (New York: CRC Press, 2019), ed. Roman V. Yampolskiy, 357-74.

optimally mitigate x-risk. We claim only that it is doubtful, in view of the available evidence and the possibility of less pathologized alternatives, whether democracies would always do so.

II. Existential Risk and Rawls' Theory of Justice

II.I. Existential Risk and the Original Position

The parties in the original position know the general facts about x-risk and democracy's pathologies. They also know that some non-democratic political systems are possibly less pathologized than democracies. None of this information is hidden from them behind the veil of ignorance: "There are no limitations on general information … there is no reason to rule out [general] facts."[33] Hence the parties know that it is doubtful whether democracies optimally mitigate x-risk.

As we have seen, the parties must agree upon principles of justice which they would want all previous generations to have followed. Since they "have no information as to which generation they belong," and since "the different temporal position of persons and generations does not in itself justify treating them differently," the parties cannot neglect the long term (which may turn out to be their own *short* term) or have any pure time preference whatsoever.[34] "[Q]uestions of social justice arise between generations as well as within them,"[35] because

[33] Rawls, *Theory*, 119.

[34] Ibid., 118, 259.

[35] Ibid., 118.

society is “a fair system of social cooperation between free and equal citizens from one generation to the next.”[36] The parties cannot ignore such questions.[37]

The only question of justice between generations which Rawls explores in any detail is finding a just savings principle which “insures that each generation receives its due from its predecessors and does its fair share” to “[maintain] just institutions and [preserve] their material base” for future generations.[38] Rawls also briefly mentions “the conservation of natural resources” and “a reasonable genetic policy” as other questions of justice between generations.[39] He does not mention x-risk, and in fact seems to suggest that humanity can expect perpetual economic and technological progress.[40]

Nonetheless, mitigating x-risk would undeniably be one of the parties’ greatest concerns. Because the life of a people is “a scheme of cooperation spread out in historical time,” every generation must “carry [its] fair share of the burden of realizing and preserving a just society.”[41] Every generation has a duty “to improve the standard of life of later generations of the least advantaged”; the scope of the difference principle extends across both space *and* time.[42] But if every generation has a duty to improve the standard of life of later generations, then every

[36] Rawls*, Justice as Fairness*, 133.

[37] Questions of justice between generations are often associated with the non-identity problem: the problem of evaluating actions which benefit or harm future people but also change which (and in some cases how many) future people will exist. For Rawls, however, the non-identity problem arguably does not arise. Because Rawls conceives of the life of a people as “a scheme of cooperation spread out in historical time,” he thinks that obligations to future generations arise *not* from obligations to specific future individuals but from the nature of social justice itself—which requires cooperation across both space and time.

[38] Rawls, *Theory*, 254-55. See also the rest of §44.

[39] Ibid., 118-19.

[40] Rawls seems to endorse Alexander Herzen and Kant’s view that later generations enjoy better fortunes than earlier ones (*Theory*, 254). He also seems to expect just societies eventually to reach a stage at which *no* saving is required for future generations because they have accumulated sufficient “real capital” (ibid., 255-6). Consequently, Rawls’ sketch of a just savings principle never addresses the possibility that future generations might be *worse* off than previous ones. But of course such a possibility exists—as Rawls’ own reference to the questions of “the conservation of natural resources” and “a reasonable genetic policy” itself presupposes.

[41] Ibid., 257.

[42] Ibid., 258.

generation *also* has a duty (*a fortiori*) to ensure that later generations are not wiped out altogether or consigned to dystopic conditions in which the value of their basic liberties is greatly diminished.

Therefore the parties cannot ignore the problem of x-risk. In fact, this problem would be among the most important in their deliberations—if not *the* most important. In an often overlooked remark, Rawls mentions a (zeroth) principle of justice *lexically prior* to the first principle "requiring that citizens' basic needs be met, at least insofar as their being met is necessary for citizens to understand and to be able fruitfully to exercise [their] rights and liberties."[43] Justice requires such a zeroth principle because the guarantees of the two principles themselves are worthless if citizens' basic needs are not reliably being met: "The realization of [citizens' fundamental] interests [in liberty] may necessitate certain social conditions and degree of fulfillment of needs and material wants."[44] Naturally, citizens' basic needs include life itself—the most basic need of all—because citizens cannot fruitfully exercise their liberties if they are dead. X-risk, however, poses a threat not just to citizens' lives but to *all* of their most basic needs, which are even more fundamental than the basic liberties themselves. Every generation will want previous generations to have done what they reasonably could to mitigate this threat. So—by Rawls' own lights—the parties must be greatly concerned to mitigate x-risk. For doing so is necessary to ensure that the most basic needs of citizens across uncountable generations are met.

What then? If the parties are greatly concerned to mitigate x-risk, and if they know that it is doubtful whether democracies optimally do so, then they have reason to reject the requirement

[43] John Rawls, *Political Liberalism: Expanded Edition* (New York: Columbia University Press, 2005), 7.
[44] Rawls, *Theory*, 476.

of EPL and to rule in multiple political systems as possibly just. The requirement of EPL threatens countless generations' lives and basic needs, because it rules out non-democratic systems which might mitigate x-risk better than democracies. Consequently, the parties have reason to reject EPL as a requirement of justice. More precisely: In a pairwise comparison between the two principles and a *modified* version of the two principles which excludes EPL from the scheme of equal basic liberties, the parties have reason to choose the modified version. [45] (Whether they *also* have reason to agree upon further modifications to the two principles is a separate question which we briefly consider in Section IV. Our argument itself neither requires nor precludes any such modifications.)

Importantly, excluding EPL from the scheme of equal basic liberties does not require the parties to exclude any other basic liberty or to reject the difference principle. The point of rejecting the requirement of EPL is not to rule in political systems which are oppressively illiberal or inegalitarian, but to rule in non-democratic systems which are *not* oppressive and which might mitigate x-risk better than democracies. It is rational for the parties to rule in such systems, because they cannot know which systems would optimally mitigate x-risk and because they have reason to doubt whether democracies would always do so. So it is rational for the parties to reject EPL as a requirement of justice.

Indeed, Rawls himself explicitly concedes something like this point. For he argues—very much in the spirit of experimentalism—that his theory of justice can (and must) rule in multiple *economic* systems:

> [M]arket institutions are common to both private-property and socialist regimes…. Which of these systems and the many intermediate forms most

[45] Rawls assumes that the parties deliberate by making pairwise comparisons of different conceptions of justice (*Theory*, 106-7).

> fully answers to the requirements of justice cannot, I think, be determined in advance. There is presumably no general answer to this question, since it depends in large part upon the traditions, institutions, and social forces of each country, and its particular historical circumstances.[46]

Rawls' point about economic systems applies no less to political systems. As we have already observed, it may be the case that which political system is most just cannot be determined in advance, and that the answer depends in large part upon the specific nature of each society (which the parties cannot know).[47] In that case, however, it is rational for the parties to rule in multiple political systems as possibly just, so as not to rule *out* any system which might work out best for some societies. The alternative, after all, is to rule out political systems which might have benefited not just some citizens but *everyone*. And that seems clearly irrational.

Hence it is a "win-win" for the parties to reject the requirement of EPL. Because the difference principle already requires whatever inequalities a society permits to be beneficial "for everyone, and in particular for the least advantaged members of society,"[48] rejecting the requirement of EPL can have only one of two outcomes: Either unequal political liberty will *not* be permitted in a given society because it is not beneficial for everyone—in which case nothing will have been lost by ruling it in as possibly just—or it *will* be permitted precisely because it accords with the difference principle by being universally beneficial. In either case, there is no downside to rejecting the requirement of EPL in favor of experimentalism.

Notably, our argument follows the exact specifications which Rawls lays out for a successful argument against an equal basic liberty like EPL. Rawls thinks that the two principles are the best conception of justice for a "well-ordered society under favorable

[46] Ibid., 2.
[47] Ibid., 134.
[48] Ibid., 13.

circumstances"—that is, a society idealized so that principles of justice are "generally complied with" and in which the "social conditions and level of satisfaction of needs and material wants" required for the fulfillment of citizens' interests in liberty have been attained.[49] Under less favorable circumstances, however, Rawls acknowledges that it may be necessary to weaken some of the first principle's requirements—including the requirement of EPL—and to move partly or entirely towards a more general conception of justice which does not prioritize the equal basic liberties.[50] Rawls acknowledges this possibility because he thinks that "the feasibility of the basic liberties depends upon circumstances."[51] Of course, his aim is still "to develop a political conception of justice" specifically for a liberal democratic regime.[52] Nevertheless, Rawls concedes that a variety of circumstances can justify restrictions of liberty—not just "historical and social contingencies" but also "the natural features of the human situation" and "the more or less permanent conditions of political life."[53]

Rawls thus lays out specific requirements for a successful argument against an equal basic liberty like EPL: Such an argument must show that the relevant inequality would be "to the benefit of those with the lesser liberty" and that it would be "accepted by the less favored in return for the greater protection of their other liberties."[54] Accordingly, we argue that rejecting the requirement of EPL is plausibly to the benefit of those potentially granted less political liberty in return for the greater protection of their other liberties—as well as their very lives. EPL

[49] Ibid., 215, 476.

[50] See ibid., 203-5, 214-8, etc. The more general conception of justice states that "[a]ll social values … are to be distributed equally unless an unequal distribution of any, or all, of these values is to everyone's advantage" (Rawls, *Theory*, 54). Thus it "imposes no restrictions on what sort of inequalities are permissible"—unlike the two principles themselves, which permit only certain kinds of inequalities (ibid., 55).

[51] Ibid., 217-8.

[52] Rawls, *Justice as Fairness*, 5.

[53] Rawls, *Theory*, 215.

[54] Ibid., 203-4.

is "subordinate to the other freedoms,"[55] which are themselves subordinate to the bare necessities of life. Consequently, *if* the politically less favored can better protect their lives and other freedoms by forfeiting their claim to political equality, it is to their benefit to do so. Plausibly, they *can* better protect their lives and other freedoms by doing so, because x-risk threatens both life and liberty and because it is doubtful whether democracies always optimally mitigate x-risk. Thus the politically less favored have at least some reason to forfeit their claim to political equality.

Importantly, our argument appeals to the very kinds of circumstances which Rawls says can justify restrictions of liberty: "natural features of the human situation" (such as the threat of x-risk) and "the more or less permanent conditions of political life" (such as democracy's pathologies). Thus our argument does exactly what Rawls says an argument of its kind must do: show that "political inequality is to the benefit of those with the lesser liberty."[56] More precisely (and weakly), it shows that embracing experimentalism—which permits but does not *require* political inequality—is plausibly to the benefit even of those who might possibly end up with less political liberty as a result.

In sum: X-risk threatens countless generations' lives and fundamental interests, and it is doubtful whether democracies optimally mitigate x-risk. So the parties have reason to reject the requirement of EPL in favor of political experimentalism.

[55] Ibid., 205.
[56] Ibid., 204.

II.II. Further Aspects of Rawls' Theory

Before turning to other objections to our argument, it is worth addressing some further relevant aspects of Rawls' theory.

First, as we have seen, Rawls thinks that the two principles (including EPL) become operative only under favorable conditions. But since the threat of x-risk is universal and permanent, conditions are arguably *never* favorable to the two principles, because even modern democratic societies face serious risks of extinction or civilizational collapse. It might therefore be thought that our argument carries no weight against the requirement of EPL, because it presupposes conditions in which Rawls himself thinks that the two principles are not operative.

Nevertheless, in our view, Rawls cannot deny that conditions are at least *sometimes* favorable to the two principles. For he clearly intends the two principles to be operative at least—and especially—in modern democratic societies like his own: "It is clear from the whole drift of Rawls's discussion … that he thinks some contemporary societies are past the point [at which the priority of liberty can be denied]."[57] As Tim Mulgan observes, Rawls' theory of justice is "tailored to his own society," which "by the time Rawls wrote his mature philosophical works … was stable and secure … a wealthy affluent liberal democracy[58]" Thus Rawls in his later work revises his theory *specifically* so that it can find "the most reasonable basis of social unity available to citizens of a modern democratic society."[59] Even in *A Theory of Justice* itself, Rawls devotes an entire chapter to a description of "an arrangement of institutions that fulfills [the

[57] Brian Barry, "John Rawls and the Priority of Liberty," *Philosophy and Public Affairs* 2, no. 3 (1973): 274-290, 279.

[58] Tim Mulgan, *Ethics for a Broken World: Imagining Philosophy After Catastrophe* (Durham, UK: Acumen, 2011), 160-1.

[59] Rawls, *Political Liberalism*, xxxix.

second principle's] requirements *within the setting of a modern state*."[60] Undeniably, then, Rawls believes that conditions are favorable to the two principles in modern democratic societies *even if* these societies have not yet effectively mitigated x-risk. (Recall that Rawls' own society was well aware of at least one x-risk—the risk of nuclear holocaust—which it had not just failed to minimize but also arguably *exacerbated*.) So our argument does carry weight against the requirement of EPL after all. For it shows that the case for that requirement is questionable even under the very conditions Rawls considers *most* favorable to the two principles.

If, however, Rawls *does* deny that conditions in modern democratic societies are favorable to the two principles, he does so at a heavy price. For if the mere presence of x-risk is enough to make conditions unfavorable to the two principles, then conditions will likely *never* be favorable to them, because the threat of x-risk is universal and permanent. Consequently, if Rawls does deny that conditions can be favorable to the two principles, he must *relegate* the two principles exclusively to ideal theory as an unachievable conception of justice. To be sure, Rawls does think that ideal theory is important—but *only* because it "presents a conception of a just society that we are to achieve if we can" and that we *can* achieve at least "in due course."[61] Indeed, Rawls' aim is to show that it is *possible* for a just and stable democratic society to exist over time even under the circumstances in which modern pluralistic democracies find themselves.[62] Hence Rawls must pay a heavy price if he concedes that the possibility of a just and stable society governed by the two principles is *unrealistically* utopian.[63] For then he must concede that the two principles should be rejected by any non-ideal (that is, any *actual*) society.

[60] Rawls, *Theory*, 228 (emphasis added).
[61] Ibid., 132, 216.
[62] See, e.g., Rawls, *Political Liberalism*, xxxix.
[63] Rawls intends his theory to be *realistically* utopian. See Rawls, *Justice as Fairness*, 4; A. John Simmons, "Ideal and Nonideal Theory," *Philosophy and Public Affairs* 38, no. 1 (2010): 5-36, 10; etc.

Second, Rawls rules out at least some probabilistic knowledge in the original position.[64] But our argument does not require the parties to have exact probabilistic knowledge about x-risk or anything else—only general (and not necessarily quantifiable) knowledge about the non-negligible cumulative risk of existential catastrophes, the pathologies of democracy, and so forth. The parties presumably know that a just political system must effectively mitigate risks of all kinds, including risks of existential catastrophes. If they know that much, and know that it is doubtful whether democracies optimally mitigate x-risk, then they know enough to have reason to reject the requirement of EPL.

Third, as we have seen, Rawls proposes principles of justice for *well-ordered* societies in which "[e]veryone is presumed to act justly and to do his part in upholding just institutions."[65] Such societies, of course, are quite different from actual societies. Accordingly, it might be thought that they would *not* be pathologized by ignorance, irrationality, or short-termism. This thought, however, would be mistaken. Besides the one idealizing assumption of strict compliance with the principles of justice, Rawls intends his theory of justice to be as realistic as possible, so that it can address the real problems facing modern pluralistic societies and not merely sidestep them. As John Simmons says,

> Rawls understands [his] ideal theory of justice as giving an account of what he comes to call a "realistic utopia," that is, the best we can realistically hope for, "taking men as they are and laws as they might be"…. We ask what could come into existence as a result of our choices, given the limits set by our moral and psychological natures and by facts

[64] Rawls says that "the veil of ignorance excludes all knowledge of likelihoods" (*Theory*, 134). Elsewhere, however, he qualifies the claim that the parties can make no "judgments of probability" (ibid., 150). In the end, Rawls is concerned chiefly to exclude probabilistic knowledge of specific "possible states of society" so that the parties can "evaluate principles solely on the basis of *general* considerations" (ibid.; emphasis added)—including considerations of x-risk.

[65] Ibid., 8.

> about social institutions and how humans can live under them. Ideal theory, then, "probes the limits of practicable political possibility."[66]

If Rawls' theory is to be realistically utopian—"taking men as they are"—then it cannot ignore democracy's pathologies *even under* the assumption of strict compliance. After all, the effect of these pathologies is not that most democratic citizens *fail to comply* with their duty to uphold just institutions, but that they comply with that duty in ignorant, irrational, and short-termist ways.[67] This effect cannot be made to disappear simply by assuming citizens' strict compliance with principles of justice, because its cause is not citizens' *noncompliance* but the limits of their "psychological natures." By Rawls' own lights, his theory cannot ignore these limits. So neither can it ignore the pathologies of democracy.

Lastly, the two principles of justice are adopted and applied in four sequential stages,[68] including a "stage of the constitutional convention" following that of the original position at which "rational delegates … guided by the two principles of justice" agree upon a just democratic constitution.[69] Such a constitution may "[limit] the scope and authority of majorities" by instituting judicial review, a bill of rights, supermajority requirements, and other constitutional devices—restricting the extent of EPL for the sake of "the greater security and extent of the other liberties."[70] It might be thought that some such constitutional devices could significantly improve x-risk mitigation. If so, then it might be wondered whether the parties can defer the problem of x-risk to the constitutional stage.

[66] Simmons, "Ideal and Nonideal Theory," 10.

[67] As Caplan, summarizing the relevant evidence on voter behavior, remarks: "Good intentions are ubiquitous in politics; what is scarce is accurate beliefs" (Caplan, *Myth of the Rational Voter*, 157). See also, *inter alia*, Timothy Feddersen, Sean Gailmard, and Alvaro Sandroni, "Moral Bias in Large Elections: Theory and Experimental Evidence," *American Political Science Review* 103 (2009): 175-92.

[68] Rawls, *Justice as Fairness*, 48.

[69] Rawls, *Theory*, 314.

[70] Ibid., 197, 201.

Of course, we do not doubt that several democratic constitutional devices could improve x-risk mitigation to some degree. Nevertheless, the range of institutional arrangements open to the parties' consideration at the constitutional stage is still significantly restricted by the two principles,[71] and there is no particular reason to think that *any* institutional arrangement within that narrow range would optimally mitigate x-risk. Actual democratic constitutions vary widely and already include various constraints on bare majority rule—and yet most current democracies do next to nothing to mitigate x-risk. Furthermore, as we argue in Section III.II, even proposed reforms specifically intended to counteract democracy's pathologies are unlikely to be very successful. Consequently, it is doubtful whether constitutional devices alone could significantly improve democracies' capacities to mitigate x-risk—let alone optimize them. So the parties still have reason to reject EPL as a requirement of justice rather than deferring the problem of x-risk to the constitutional stage.[72]

Again, we do not claim to *know* whether the parties should reject the requirement of EPL all things considered. But we *do* claim that the considerations which we have put forward in favor of doing so will not be easily outweighed. X-risk threatens the fundamental interests of both present and future generations. If some non-democratic systems might mitigate x-risk better

[71] Delegates in the constitutional convention can restrict the extent to which the constitution is purely majoritarian (*Theory*, 197), but only in ways which "fall equally upon everyone" (ibid., 201). They cannot, for example, violate "the precept one man one vote" (ibid.).

[72] It might also be wondered whether the parties can defer the problem of x-risk to the *legislative* stage "in which laws are enacted as the constitution allows and as the principles of justice require and permit" (*Justice as Fairness*, 48). Maybe the parties can solve the problem of x-risk simply by ensuring that democratically elected legislative bodies enact laws which effectively mitigate it. But if Rawls' theory is to be realistically utopian, it must account for the general facts about democracy in virtue of which it is *doubtful* whether democratically elected legislative bodies can be expected to enact such laws. Deferring the problem of x-risk to the legislative stage simply begs the question against our argument by ignoring the relevant general facts and the possibility that some non-democratic political systems might mitigate x-risk better than democracies.

than democracies, then the parties have reason to rule in such systems by rejecting the requirement of EPL.

III. Objections and Replies

Two objections to our argument merit further discussion. The first—*the objection from epistemic democracy*—is that voter ignorance does not significantly compromise democratic decision-making, and that democracies can actually *outperform* other political systems by drawing upon the collective intelligence of crowds. The second—*the objection from democratic reform*—is that suitably reformed democracies *would* optimally mitigate x-risk even if actual democracies do not. If either of these objections succeeds, then it may no longer be doubtful whether democracies optimally mitigate x-risk. And then our argument fails.

III.I. The Objection from Epistemic Democracy

We have claimed that typical voters are ignorant of basic political and other facts, and that politicians are incentivized to respond to voters' ignorant (and irrational) preferences. Partly on the basis of these claims, we have argued that democratic decision-making is compromised both generally and specifically with respect to x-risk. Epistemic democrats, however, resist such claims, and argue that agents who are individually ignorant can still make collectively intelligent decisions under the right conditions.[73]

[73] Hélène Landemore, *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many* (Princeton, NJ: Princeton University Press, 2013); Melissa Schwartzberg, "Epistemic Democracy and Its Challenges," *Annual*

Some epistemic democrats, appealing to Condorcet's jury theorem, argue that democratic collectives can make intelligent decisions so long as their members vote sincerely and independently with a probability greater than 0.5 of voting for the "correct" outcome.[74] Some argue that larger and more diverse decision-making groups can epistemically outperform smaller and less diverse groups, even if the latter are made up of experts.[75] Since the larger and more diverse groups can draw upon the distributed knowledge of their members more effectively, they can collectively know more than the smaller groups of experts (even if every expert knows more than every non-expert). Less ambitiously, others argue that voters can use a variety of heuristics to overcome their political ignorance by learning from political parties, opinion leaders, traditional and online media, and other sources.[76] By using simple heuristics, uninformed voters can make competent political decisions by tracking the beliefs of others who are better-informed. If such heuristics are reliable, then voter ignorance need not significantly compromise democratic decision-making.

If epistemic democrats' claims are correct, then we have overstated the extent to which voter ignorance pathologizes democracy. It seems to us, however, that such claims on behalf of democracy are overly optimistic.

First, voting in the real world is nothing like voting in accordance with Condorcet's jury theorem, which requires not only voter independence and sincerity but also a minimum voter competence higher than is warranted by the available evidence. It is likely false that enough

Review of Political Science 18 (2015): 187-203; Kai Spiekermann and Robert E. Goodin, *An Epistemic Theory of Democracy* (Oxford: Oxford University Press, 2018).

[74] For further discussion, see Landemore, *Democratic Reason*, 147-56.

[75] Scott E. Page, *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton, NJ: Princeton University Press, 2007); Landemore, *Democratic Reason.*

[76] Samuel L. Popkin, *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns* (Chicago: University of Chicago Press, 1991).

voters have a probability greater than 0.5 of voting for the correct outcome with respect to many important political issues, including x-risk.[77] Indeed, some evidence suggests that voters are systematically *mistaken* about many important political and economic issues.[78] Even if they are not, the correct outcome with respect to a given political issue is often not on the ballot. And in most cases *no* outcome addressing x-risk is on the ballot whatsoever.

Second, although *some* larger and more diverse groups can epistemically outperform smaller and less diverse groups, not all such larger groups can. A minimum threshold of competence among members of larger groups is necessary for them to outperform smaller groups epistemically,[79] and the relevant evidence from political science appears to show that democracies do not meet this threshold. In theory, of course, democracies can draw upon the wisdom of crowds.[80] In practice, however, they often do not. And even if they always did, some smaller and less democratic groups of experts might still epistemically outperform them and thus better mitigate x-risk.

Lastly, the appeal to reliable heuristics faces several serious problems. Most obviously, many heuristics used by voters are clearly *unreliable*. Political parties often pander to voters' misconceptions rather than reliably tracking the truth, and both traditional and online media not only fail to report facts reliably but often spread misinformation.[81] Furthermore, even when voters do have access to reliable heuristics, determining *which* heuristics are reliable is itself a

[77] For critical discussion of attempts to apply Condorcet's jury theorem to actual democracies, see Brennan, *Against Democracy*, 179-80. Although some of the theorem's assumptions can be relaxed (Goodin and Spiekermann, *Epistemic Theory of Democracy*, 17-83), even weakened versions of Condorcet's jury theorem require *some* threshold of minimum voter competence, and we doubt whether actual democracies meet any such =threshold.

[78] Caplan, *Myth of the Rational Voter*.

[79] Brennan, *Against Democracy*, 182-85.

[80] James Surowiecki, *The Wisdom of Crowds* (New York: Anchor Books, 2004).

[81] Yochai Benkler, Robert Faris, and Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (Oxford: Oxford University Press, 2018).

costly and cognitively demanding task—which most voters rationally refrain from undertaking just as they rationally refrain from becoming politically well-informed in general. In fact, models of rational irrationality suggest that voters often choose heuristics for reasons unrelated to their reliability, such as their entertainment value or congruence with voters' preexisting views.[82] Thus the very same ignorance and irrationality which keep voters from forming rational beliefs also keep them from identifying and using reliable heuristics.

But maybe epistemic democrats need only appeal to comparisons of current democracies with current non-democracies to show that democracies *do* outperform other political systems on several important fronts—including many related to long-term, complex problems like x-risk. It is sometimes argued, for instance, that democracies deal with climate change better than non-democracies.[83] And democracies are also wealthier than non-democracies, much less prone to famine, less warlike, and so on.[84] Arguably, then, current democracies outperform current non-democracies *in general*. So perhaps we should also expect them to outperform other systems with respect to x-risk.

Nonetheless, for at least two reasons, we doubt whether current democracies' superior performance to current non-democracies significantly undercuts our argument. First, although a thorough examination of the relevant evidence is beyond the scope of this essay, it is worth noting that attributions of democracies' superior performance *specifically* to their democratic institutions are often controversial. Some argue that democracies are generally more prosperous

[82] Somin, *Democracy and Political Ignorance*, 90-118.

[83] Daniel J. Fiorino, *Can Democracy Handle Climate Change?* (Medford, MA: Polity Press, 2018).

[84] Daron Acemoglu, Suresh Naidu, Pascual Restrepo, James A. Robinson, "Democracy Does Cause Growth," *Journal of Political Economy* 127 (2019): 47-100; Amartya Sen, "Democracy as a Universal Value," *Journal of Democracy* 10 (1999): 3-17; R.J. Rummel, "Democracies Are Less Warlike Than Other Regimes," *European Journal of International Relations* 1 (1995): 457-79.

and peaceful not because of their *democratic* institutions but because of their broadly *liberal* ones.[85] In a similar vein, Stefan Wurster argues that democracies' superior performance with respect to climate change is at best restricted to adaptations to "area-restricted environmental problems and those that are technically easy to solve."[86] As we have seen, however, x-risks are not problems with easy solutions or ones to which we can gradually adapt. Once an existential catastrophe occurs, no recovery is possible.

Second, and even more importantly, current democracies' superiority to current non-democracies hardly entails their superiority to *all possible* non-democratic systems. Even if current democracies' relative success *is* partly caused by their democratic institutions, the pathologies of democracy still hinder their capacities to deal with long-term, complex problems like x-risk. It thus remains possible that democratic x-risk mitigation is *suboptimal* relative to that of some non-democratic systems. Since the parties cannot rule out this possibility, they still have reason to reject EPL as a requirement of justice.

Therefore the objection from epistemic democracy fails. The parties in the original position continue to have reason to reject the requirement of EPL, because democracies' capacities to mitigate x-risk remain doubtful.

[85] John R. Oneal and Bruce Russett, "Assessing the Liberal Peace with Alternative Specifications: Trade Still Reduces Conflict," *Journal of Peace Research* 4 (1999): 423-42; Håvard Hegre, "Democracy and Armed Conflict," *Journal of Peace Research* 2 (2014): 159-72. For further discussion, see Garett Jones, *10% Less Democracy: Why You Should Trust the Elites a Little More and the Masses a Little Less* (Stanford, CA: Stanford University Press, 2020), 15-17.

[86] Stefan Wurster, "Comparing Ecological Sustainability in Autocracies and Democracies," *Contemporary Politics* 19, no. 1 (2013): 76-93.

III.II. The Objection from Democratic Reform

Even if current democracies do not optimally mitigate x-risk, it might be argued that *suitably reformed* democracies would do so. Maybe the right democratic reforms could counteract democracy's short-termism and other pathologies. If so, then the parties arguably no longer have reason to reject the requirement of EPL, since they can ensure effective x-risk mitigation simply through democratic reform.

Democratic theorists have proposed several reforms to counteract democratic short-termism. First, constitutions could be amended to include provisions safeguarding future generations' interests.[87] Among other things, such provisions could safeguard those interests by penalizing policymakers who violate them.[88] Second, and in conjunction with the first proposal, ombudsmen could be selected to ensure that policymakers do not violate constitutional provisions safeguarding future generations' interests.[89] Third, quotas could be imposed on legislative bodies requiring a certain minimum proportion of younger representatives, with the expectation that these representatives would prioritize long-term issues more.[90] Fourth, voting could be weighted by *age*, so that younger citizens' votes would be weighted more heavily than those of older citizens.[91] Assuming that legislators were responsive to the political preferences of

[87] Ernst Brandl and Hartwin Bungert, "Constitutional Entrenchment of Environmental Protection: A Comparative Analysis of Experiences Abroad," *Harvard Environmental Law Review* 16 (1992): 1-100; Kristian Skagen Ekeli, "Green Constitutionalism: The Constitutional Protection of Future Generations," *Ratio Juris* 20, no. 3 (2007): 378-401; Iñigo González-Ricoy, "Intergenerational Provisions," in *Institutions for Future Generations*, eds. González-Ricoy and Gosseries.

[88] González-Ricoy, "Intergenerational Provisions," 170.

[89] Ludvig Beckman and Fredrik Uggla, "An Ombudsman for Future Generations: Legitimate and Effective?", in *Institutions for Future Generations*, eds. González-Ricoy and Gosseries.

[90] Juliana Bidadanure, "Youth Quotas, Diversity, and Long-Termism: Can Young People Act as Proxies for Future Generations?", in *Institutions for Future Generations*, eds. González-Ricoy and Gosseries.

[91] Philippe van Parijs, "The Disenfranchisement of the Elderly, and Other Attempts to Secure Intergenerational Justice," *Philosophy and Public Affairs* 27, no. 4 (1998): 292-333. More radically, citizens over a certain age could be disenfranchised altogether (ibid.). Such a reform, however, would scarcely count as *democratic*.

the young (and that those preferences were in fact more long-termist), age-weighted voting could make democracies more long-termist by increasing younger voters' political influence. Fifth, legislative bodies could be set up whose members were selected at random from the general population.[92] Free from short-termist electoral pressures, and guided by expert advice, such bodies could effectively counterbalance more short-termist electoral bodies. Sixth, and relatedly, legislative bodies could be set up with specific mandates to represent the interests of the young and (by extension) future generations.[93]

For our part, we do not deny that reforms like these could counteract democratic short-termism to some degree. Nonetheless, we doubt whether such reforms would *optimize* democratic x-risk mitigation, since they fail to account sufficiently for the impact of voter ignorance and irrationality on democratic decision-making. This failure is perhaps unsurprising, because almost every proposal in the literature on long-termist reforms *presupposes* democracy's necessity for long-termist politics. In our view, however, such a presupposition is ill-advised. After all, the range of possible political systems is immense, and there is no reason to set aside all non-democratic systems *in advance* without first considering their capacities to mitigate x-risk. This is especially so because the risk-mitigating capacities *even* of reformed democracies are themselves doubtful.

Consider first a democracy reformed in accordance with the first and second proposals listed above, so that its constitution is amended to include provisions safeguarding future

[92] Michael K. MacKenzie, "A General-Purpose, Randomly Selected Chamber," in *Institutions for Future Generations*, eds. González-Ricoy and Gosseries; Smith, *Can Democracy Safeguard the Future?*

[93] Tyler M. John, "Empowering Future People By Empowering the Young?, in *Ageing Without Ageism: Conceptual Puzzles and Policy Proposals*, eds. Greg Bognar and Axel Gosseries (forthcoming). Additionally, iterated mechanisms of retrospective accountability could be instituted to reward (or punish) policymakers for the long-term effects of their policies (ibid., 17-8).

generations' interests and ombudsmen for future generations are selected. Even if such a democracy would be more long-termist than actual democracies, we do not yet have any reason to think that it would deal with x-risk (and other long-term problems) *optimally*. Without further measures to counteract political ignorance and irrationality, such a democracy could do little more than replace ignorant and irrational *short*-termism with ignorant and irrational *long*-termism. No doubt, a greater regard for the future is *necessary* to mitigate x-risk—but it is hardly therefore *sufficient*. For our goal is not just that political systems *try* to mitigate x-risk (though most existing systems fail to do even that) but that they mitigate x-risk *well*.

Second, consider a democracy reformed in accordance with the third and fourth proposals, so that youth quotas are imposed on its legislative bodies and younger citizens' votes are weighted more heavily. Even if we grant the assumption that younger citizens' political preferences are more long-termist than those of older citizens, the point still remains that a greater regard for the future is not sufficient for effective long-termist policy. Whether or not younger citizens' political preferences are in fact more long-termist, it is the *quality* of those preferences that matters most. Since most younger citizens (like most citizens in general) are ignorant and irrational, increasing their political influence hardly guarantees better long-term political outcomes. Moreover, in any case, both youth quotas and age-weighted voting clearly *violate* the requirement of EPL (which guarantees equal access to public office and one vote per citizen). So the possible effectiveness of these reforms scarcely counts against our argument that the parties have reason to *reject* that requirement.

Lastly, consider a democracy reformed in accordance with the fifth and sixth proposals, so that legislative bodies are set up whose members are randomly selected from the general population and which have specific mandates to represent future generations' interests. Of the six proposals listed above, we suspect that these two would best mitigate voter ignorance and irrationality, since a democracy reformed in accordance with them could enable selected citizens to become better informed by exposing them to expert feedback and sustained and focused deliberation.[94] Nonetheless, at least two problems with these proposals remain. First, the legislative bodies they call for would have only limited, chiefly advisory powers. Second, and more importantly, such bodies would not necessarily outperform *other* bodies whose members met more demanding selection requirements. It is entirely possible that legislative bodies which screened out especially ignorant and irrational citizens would outperform those which failed to do so—both generally and specifically with respect to x-risk. So it is doubtful whether democracies reformed in accordance with the fifth and sixth proposals listed above would mitigate x-risk optimally.

Of course, our list of proposed long-termist democratic reforms is far from exhaustive, and we cannot assess every such proposed reform here.[95] Nevertheless, our discussion of the six proposals listed above reveals a common recurring flaw in such proposals: a failure to account sufficiently for the impact of voter ignorance and irrationality on democratic decision-making. This recurring flaw makes it doubtful whether democracies reformed in accordance with such proposals would optimally mitigate x-risk. Accordingly, the parties still have reason to reject the requirement of EPL.

[94] James S. Fishkin, *When the People Speak: Deliberative Democracy and Public Consultation* (Oxford: Oxford University Press, 2009), 106-58.

[95] For a more thorough survey of proposed long-termist democratic reforms, see *Institutions for Future Generations*, eds. González-Ricoy and Gosseries.

IV. Conclusion

We have argued that democracy is pathologized by voter ignorance, voter irrationality, and short-termism. Since it is possible that some other political systems are not comparably pathologized, it is doubtful whether democracies optimally mitigate x-risk. But the parties in the original position are greatly concerned to mitigate such risk. So the parties have reason to reject the requirement of EPL in favor of political experimentalism.

We have focused on the deliberations of the parties in the original position because the original position is the centerpiece of Rawls' theory of justice. Nevertheless, the thrust of our argument is not restricted by this focus. To be sure, the original position is no more than a "device of representation" which is always subject to possible revision.[96] But our argument does not flow narrowly from Rawls' description of the original position itself. Instead, it flows broadly from some of Rawls' foundational normative commitments (along with the relevant general facts): a commitment to the fundamental importance of citizens' lives and basic needs to their interests; a commitment to the equal weight of the interests of all generations; and so on. Our challenge to Rawls is not to revise these commitments but to incorporate the relevant general facts about x-risk and democracy into his theory. These facts give Rawls reason to reject EPL as a requirement of justice.

Our argument may also give Rawls reason to make other changes to his theory besides rejecting the requirement of EPL. For example, if other equal basic liberties hinder x-risk mitigation, Rawls may have reason to reject these other basic liberties as requirements of justice.

[96] Rawls, *Theory*, 17.

Furthermore, since x-risks themselves are not the only threats to citizens' fundamental interests, *other* risks—such as risks of war or climate change—may also prompt changes to his theory.[97]

More importantly, our argument has far-reaching implications beyond Rawls' theory itself. At its core, our argument flows from two plausible normative commitments which are shared by Rawlsians and non-Rawlsians alike: a commitment to the fundamental importance of people's lives and basic needs to their interests, and a commitment to the non-negligible (if not equal) weight of the interests of future generations. Neither of these commitments is uniquely Rawlsian; in fact, almost all of us share them. And because we do, we can agree that political systems must safeguard future generations' fundamental interests by effectively mitigating x-risk. It is doubtful, however, whether democracies optimally mitigate x-risk. So almost all of us have at least some reason to reject the requirement of EPL in favor of experimentalism.

In fact, we may have even further reason to reject the requirement of EPL, since there may be other threats to our fundamental interests which democracies might not optimally mitigate. Though x-risk is the most sweeping such threat, it is far from the only one, and it is possible that other such threats may give us reason to reject EPL as a requirement of justice. We cannot discount this possibility, because democracy's pathologies compromise democratic decision-making *in general*—not only with respect to x-risk. Consequently, our argument can be generalized in at least two important ways: first, beyond Rawls himself to anyone else with similar normative commitments; second, beyond x-risk itself to any other threat to present or future generations' fundamental interests which democracies might not optimally mitigate.

[97] For further discussion along similar lines, see Mulgan, *Ethics for a Broken World*, 160-96.

It is hard to deny that the loss of humanity's long-term potential would be an unfathomable tragedy, erasing countless generations and trillions of lives. The reason that mitigating x-risk should be one of the parties' greatest concerns is that it should be one of *our* greatest concerns. If it is not, then we can only conclude that ordinary human psychology has things precisely backwards—for a single death is a tragedy; but so are a trillion.