

DEMOCRACY AND ANTHROPIC RISK

Petr Špecián

Charles University

Prague University of Economics and Business

petr.specian@fhs.cuni.cz

Abstract

Democracy in its currently dominant liberal form has proven supportive of unprecedented human flourishing. However, it also appears increasingly plagued by political polarization, strained to cope with the digitalization of the political discourse, and threatened by authoritarian backlash. A growing sense of the anthropic risks—with runaway climate change as the leading example—thus often elicits concern regarding democracy’s capability of mitigating them. Apparently, lacking a sufficient degree of the citizens’ consensus on the priority issues of the day, it can find itself unable to muster the resolve and resources necessary to alter its “business as usual” path even where disasters loom ever larger in its track. Against these concerns, I argue that democracy’s potential remains far from depleted. While political irrationality needs to be considered a major obstacle to successfully managing the anthropic risks, many existing reform proposals appear capable of reining it in. My main suggestion is to seek ways to facilitate a speedier and more effective exploration of the landscape of the possible democratic mechanisms of collective choice. A virtual platform to support such an exploration — which I call ‘democracy’s digital playground’ — may allow us to find innovative ways of managing anthropic risks without compromising our commitment to democratic values.

Keywords: Anthropic risk; Democratic reform; Institutional design; Political Irrationality.

Resumo

A democracia na sua forma liberal atualmente dominante demonstrou suportar um florescimento humano sem precedentes. No entanto, também parece cada vez mais atormentada pela polarização política, forçada a lidar com a digitalização do discurso político e ameaçada pela reação autoritária. Uma crescente consciencialização dos riscos antropogénicos — com a mudança climática descontrolada como principal exemplo — muitas vezes levanta preocupações sobre a capacidade de a democracia mitigá-los. Aparentemente, sem um grau suficiente de consenso dos cidadãos sobre as questões prioritárias do dia, pode ser incapaz de reunir a determinação e os recursos necessários para alterar o padrão “business as usual”, mesmo quando os desastres se apresentam cada vez maiores mantendo o seu curso atual. Diante dessas preocupações, defendo que o potencial da democracia continua longe de estar esgotado. Embora a irracionalidade política deva ser considerada um grande obstáculo para a gestão bem-sucedida dos riscos antropogénicos, muitas propostas de reforma existentes parecem capazes de controlá-los. A minha principal sugestão é procurar modos de facilitar uma mais rápida e mais eficaz exploração do panorama dos possíveis mecanismos democráticos de escolha coletiva. Uma plataforma virtual para apoiar essa exploração — que denomino “playground digital da democracia” — poderia permitir-nos encontrar formas inovadoras de gerir riscos antropogénicos sem comprometer o nosso compromisso com os valores democráticos.

Palavras-chave: Risco antrópico; Reforma democrática; Desenho institucional; Irracionalidade política.

Introduction

Human capabilities are growing at a breathtaking, exponential speed. There seems to be no limit to the aspirations of the global civilization. Concurrently, however, unprecedented risks have emerged, driven by the growing scale and magnitude of human activities. Among these *anthropic risks*, climate change has become perceived as the most urgent one. But it sure is not the only challenge humanity faces. Consider the thousands of nuclear warheads on their hair-trigger or the recently demonstrated unstopability of pandemics in our deeply interconnected world. And sadly, other threats loom on the horizon. Powerful new technologies, such as synthetic biology or artificial intelligence, have a frightening potential downside (Ord, 2020). While not all of them may live up to their current hype, there shall be others to take their place. Even in the best-case scenarios, the manageability of anthropic risks remains an open question.

Technologies increase the individuals' ability to shape their environment. They can be utilized for the common benefit or as a tool of destruction. However, it is the collective decision-making – and its outcomes in terms of institutions and incentives – that largely determines which technologies will be developed and to which uses they shall be put. The functioning of the mechanisms of collective choice is thus crucial for the sustainability of human civilization. When they break down, perhaps due to an epistemic crisis (Špecián, 2022b), a society becomes unable to steer away from its trajectory even in the face of an impending disaster.

Among the modes of governance that people have stumbled upon so far, liberal democracy appears the most capable of facilitating their flourishing. However, this guarantees neither its continuing resilience nor its overall optimality. Today, there is a growing sense that liberal democracy finds itself in a crisis (e.g., Repucci, 2021). Its ability to cope with the centrifugal forces of political polarization may have become curtailed by the increasing decentralization of ways in which information, as well as disinformation, is created and disseminated in society, that is, by “epistemic democratization” (Sismondo, 2017).

Against this background, I shall defend two claims. First, even in the face of anthropic risks, democracy's potential remains far from depleted. Second, the key challenge for humanity – if it is to sustain its civilization – is to unleash the exploration of democracy's design space. There are many possible institutional embodiments of democratic principles. The task is finding such that will prove capable of coping with the challenges of the day.

1. Democracy's Achievements and Potential

Democracy is a concept contested so deeply that there is no chance of finding any consensual definition. Nonetheless, for the sake of the current exercise, let me use “democracy” to denote any mechanism of collective choice that is *inclusive* and *avoids exogenous value imposition*. Here, inclusivity means

that all of the society's members have the right to express their beliefs, interests, and – ultimately – preferences in the course of the collective decision-making. Such a right exists by default and can only be abrogated based on a narrow set of well-defined limitations, such as a minimum age requirement. Avoiding exogenous value imposition (or *non-imposition*, for short) means that the preferences of the society's members also determine the decision made. In other words, a democratic mechanism is only such that allows the people themselves to decide and prevents third parties from dictating their will. While a degree of value imposition is often necessary since consensus proves too elusive, it is only consistent with the democratic way of decision-making if it is endogenous, such as in the case of majority choice where minority preferences end up being overruled.

Can a mechanism of collective choice, which is effective enough to manage anthropic risk, be democratic? There exist strong theoretical and empirical reasons for optimism in this regard. As once pointed out by Hayek (1945), the key challenge to any complex society is information processing. Valuable knowledge is dispersed among the many members of a society, and a centralized decision-making body finds itself strained to access but a tiny sliver of it. Its ignorance then makes it prone to inefficiencies and – possibly catastrophic – misjudgments.

While Hayek's argument favors markets as the institutional mechanism most capable of aggregating the dispersed knowledge, they are not the only game in town. When functioning correctly, which is unfortunately not always the case, they represent a case of a broader phenomenon known as *the wisdom of crowds* (Surowiecki, 2005). The wisdom of crowds emerges when large numbers of individually unsophisticated decision-makers solve complex epistemic tasks with performance often exceeding that of the experts. This happens in well-functioning markets, and it can also be achieved in well-functioning democracies (Landemore, 2012). Like the market, democracy – with its inclusivity and non-imposition – also allows harnessing the dispersed knowledge of the many.

As the currently dominant embodiments of democratic principles, liberal democracies provide a persuasive demonstration of democracy's potential. They outperform their autocratic counterparts across a range of basic measures, such as peacefulness, prosperity, or protection of basic rights (e.g., Christiano, 2011). While their critics may muster much evidence to complain about regular citizens' lack of knowledge, interest, and even competence in elementary cognitive tasks (Brennan, 2016), they find themselves strained to point out a non-hypothetical political regime that represents an attractive alternative. It is telling that democracy's appeal proves so irresistible that even the most brutal dictatorships find themselves compelled to pay lip service to its core values.

Still, liberal democracy – however marvelous its success story – is likely not the most stable democratic system in the stormy waters of the 21st century,

let alone the best possible one overall. It represents a particular equilibrium in democracy's design space: one among the many institutional setups that conform to the requirements of inclusivity and non-imposition. Sure, it has proven resilient in many past crises and provides a decent payoff. However, there may be a dominant equilibrium elsewhere, waiting for discovery. Also, historical robustness is no guarantee of future robustness. The swift development of the technological substrate of our civilization may render liberal democracy more fragile and less effective. Many fear that the digital revolution, which is reshaping democracy's underlying infosphere, dramatically endangers its prospects (e.g., Gurri, 2018).

2. Beware Political Irrationality

Clearly, humans are not infallible. We make mistakes, both as individuals and as collectives. As it appears, the magic of the wisdom of crowds is conditional on its ability to suppress the influence of mistakes by aggregating them out. If one person randomly misjudges a situation, the mistake may obscure any hints of useful information her guess contains. If a hundred people commit a random misjudgment, their mistakes – even if individually significant – shall average out, cleansing the signal. However, such a fortuitous impact of aggregation is conditional on several premises, such as the mistakes being random and the individuals sincerely trying to get things right (cf. Goodin & Spiekermann, 2018). Unfortunately, neither of these premises can be taken for granted: we, the *homo sapiens*, are prone to systematic mistakes on account of our cognitive quirks and limitations, as well as to insincerity triggered by strategic considerations. In short, people can be both too irrational and too rational for the wisdom of crowds to emerge.

Political irrationality means that people systematically choose political alternatives they do not most prefer. How could anything so paradoxical even occur? There seem to be two paths headed in such a direction. The first is bounded rationality (Kahneman, 2003). Since our cognition is imperfect, we are prone to a predictably biased assessment of the available information in various circumstances. For instance, we tend to find information more credible if it confirms our baseline beliefs than when it counters them. This tendency – a *confirmation bias* – prevents our factual beliefs from converging to the truth as more evidence arrives. A salient example of the lack of truth-convergence is the persistent absence of social consensus on the existence and causes of climate change. The second avenue is *motivated reasoning* (Epley & Gilovich, 2016). Humans are social creatures: our motivations are not merely cognitive. Sometimes, getting along with others is more important than sincerely professing one's genuine views. In the context of mass democracy, where a single vote's probability of being decisive is very close to zero, there arises a temptation to use one's voice to demonstrate one's allegiance to an in-group

while ignoring the substance of the matter itself. Thus, one may ultimately vote against an alternative that, if taken on its own terms, one prefers and would choose if he or she were decisive.

Worryingly, the digitalization of the democratic discourse may exacerbate both issues, increasing the probability that political polarization will tear a democratic society apart. A voluminous literature blames the novel ways of (dis)information dissemination – especially social media, such as Twitter and Facebook – for liberal democracy’s current plight (e.g., Sunstein, 2017). These media allow for cheap and accurate matching between content producers and consumers with their algorithms always seeking novel ways to increase user engagement. Unfortunately, such a business model has unfortunate side effects in the context of both bounded rationality and motivated reasoning (Špecián, 2022a, Chapter 3).

Consider citizens’ bounded rationality. There persists a controversy regarding the degree to which it is “hardwired,” that is, impervious to learning or change of incentives. Still, even if bounded rationality proves largely curable – once people become more acquainted with the specifics of the digital discourse, for instance – as far as it persists, it makes one vulnerable to exploitation. Bounded rationality leads to systematic, predictable mistakes, and anyone aware of their existence may attempt to trigger them on purpose. Since the digital environment enables precise audience targeting, rapid and reliable feedback on a message’s impact, and efficient quasi-evolutionary selection of potent memes, it represents a perfect niche for the bad agents to benefit from their victims’ vulnerabilities. Arguably, the ascent of *fake news* provides an example of the digitalization’s repercussions (Gelfert, 2018).

With motivated reasoning as the culprit of political irrationality, the situation is hardly any rosier. The environment of social media motivates the use of political statements as signals of one’s identity and group allegiance more strongly than offline interaction. Any statement is recorded and searchable. Compartmentalization between different areas of life is all but impossible. One can no longer hope that a silly comment will eventually be forgotten or that she can wear a different political coat when interacting with a different group of people. While this may sound alarming, there also arises an excellent opportunity to signal one’s ideological commitments *credibly*: a person who conspicuously burns the bridges to the other side, making herself permanently “unclubbable” to the competing out-groups, can be trusted as a true ally (Mercier, 2020; Simler & Hanson, 2018). At least part of the fake news’ success is thus explainable as a manifestation of motivated reasoning (Kahan, 2017).

3. Democratic Reform: An Embarrassment of Riches

How to better fortify democracy against political irrationality? There is no paucity of proposals. Some of them directly target the misinformation epidemic

of late, distinguishing between “educative approaches” that teach people how to become more effective critical reasoners in the digital environment and “structural approaches” that strive to alter the environment itself to make it less challenging, given people’s constrained resources and cognitive faculties (Croce & Piazza, 2021). Others address irrationality more broadly. It is perhaps convenient to distinguish between *marginal reforms* that strive to disarm one bias at a time by targeted interventions and *radical reforms* that do not shy away from fundamentally rethinking mechanisms of representation and preference revelation in the context of collective choice (Špecián, 2022a, Chapter 5).

Among the marginal reforms, *nudges* (Thaler & Sunstein, 2008) have received the broadest acclaim in academia and beyond. These non-intrusive and non-coercive alterations to the design of the environment in which the individuals make their choices are supposed to provide gentle guidance towards decisions that are better for the choosers “as judged by themselves.” Unfortunately, their proper application is conditional on first identifying the individuals’ underlying, mistake-free preferences, which appears to be a rather tricky task (Špecián, 2019). However, there are other options. *Boosts* seek to empower individuals by enhancing their competencies, such as the ability of statistical reasoning (Hertwig & Grüne-Yanoff, 2017). *Budges*, in contrast, are regulations that strive to prevent the use of exploitative communication techniques by suppliers in commercial – and perhaps also political – marketplaces (Oliver, 2013).

Prominent radical reform proposals include *open democracy*, which rethinks political representation (Landemore, 2020), or *quadratic voting*, which transforms how preferences are revealed in collective choice (Posner & Weyl, 2018, Chapter 2). With open democracy, elected representatives get replaced by a random sample of the population. The lottocratic representatives then participate in deliberative sessions before taking a majority vote on the issues in their purview. Political power thus transfers from professional politicians, who are perhaps too easily corruptible or too mired in the issues of merely symbolic importance, to the people themselves. With quadratic voting, a hallowed one-person-one-vote principle of democratic collective choice is jettisoned for the sake of a system that enables the recording of preferences’ intensity. It is thus no longer the case that a tepid preference in one direction provides a perfect counterweight to an intense preference in the opposite direction when it comes to counting votes. Quadratic voting enables each citizen to allocate scarce voice credits to various alternatives based on how strongly he feels about them. At the same time, the quadratic nature of the procedure, which counts the votes as a square root of the expended voice credits, prevents the most extreme preferences from capturing the election.

The above examples are hardly exhausting. There are dozens, if not hundreds, of reforms that have been proposed and show a theoretical promise of improving democracy’s performance vis-à-vis anthropic risks: a true

embarrassment of riches. However, these proposals largely remain on paper only. Even those most influential and thoroughly tested – such as James Fishkin’s deliberative polls (e.g., Fishkin, Siu, Diamond, & Bradburn, 2021) – have made little tangible difference in the actual political outcomes. Uncertainty lingers regarding how they would perform in the political wilderness beyond the neat confines of an experimental setting. As it appears, the problem is not as much designing an innovative institutional mechanism as applying and testing it in realistic circumstances. And even more is needed: since we must avoid the false presumption that a society can be reconstructed according to a blueprint, we need not only a space to test the designs engineered by scholars but also a space for novel ways of democratic decision-making to evolve and grow organically.

4. Democracy’s Digital Playground

The institutional design space is vast, but human efforts to explore it have been limited, so far. Consider how static and conservative our institutions are compared to our physical technologies. The technological change since the Industrial Revolution has been much more profound than the institutional change. In fact, the last centuries have witnessed a significant loss in institutional diversity triggered by a broad convergence towards the Western institutional model.

Of course, institutional conservatism has its benefits: stability and continuity should not be discounted as important contributors to human flourishing. The same holds for emulation of effective institutional designs: why reinvent the wheel if you can just copy the one your neighbor uses with much success. Yet, the looming anthropic risks should perhaps prod us to more urgently consider the steps that could be taken toward unleashing human ingenuity for the sake of exploration of the institutional design space.

Let me venture a speculative proposal in the given direction: namely, a call for instituting *democracy’s digital playground*. Where so many authors busy themselves constructing novel schemes of democratic reform, democracy’s digital playground might qualify as a meta-reform. Its purpose is to provide a space where democratic institutional mechanisms could compete, driving innovation forward: the unviable ones would be weeded out, the viable ones tweaked and further improved due to an unrelenting pressure of selection. How might this work?

The innovation space would be delimited by minimum consensus democratic values, that is, by basic requirements on democraticity that command broad public support and align with what democracy means to the regular people. The principles of inclusivity and non-imposition appear quite attractive, for instance. So, any institutional mechanism installed on the playground would be required to allow meaningful participation of its users in the

collective choices directing its operation. Within the confines of the minimum consensus values, however, free exploration of the design space would be allowed.

The space itself must be digital rather than physical. Historically, institutional innovation has been hampered by the need to find a geographic location not already controlled by the existing institutions. Such a process is far too costly to allow a sufficiently dynamic institutional innovation. Digital space, on the other hand, is where real estate is cheap and moving between different jurisdictions – that is, voting by one’s feet – can be practiced with relative ease. Besides being more fluid, it is also less burdened by the heavy hand of the existing institutional mechanisms. Just consider how digital migration is much simpler than physical migration, especially with the increasing accessibility of the internet. Digital space thus offers an enhanced opportunity for democratic participation even for the poor and disenfranchised of the world who otherwise possess little hope of physically escaping their circumstances.

Finally, the space must be a playground. One good reason our institutions are so rigid is that any major overhaul of the status quo is immensely dangerous. What if you replace representative democracy with open democracy, for instance, and the whole political order crumbles? But for learning, trial and error are of the essence. If we are to encourage innovation, failures must be survivable. That is why children’s playgrounds are such important places for one’s development: the stakes are real – you can still get a bloody nose or knee – but the environment is set up to keep the possible harms within reasonable limits. Mistakes are punished, but not so severely that it would be hard to recover and retry.

Currently, the most elaborate digital worlds that come closest to democracy’s digital playground as I conceive of it are to be found in massively multiplayer online games. Of course, these games come short of an ideal template in many ways. For instance, while it is easy for the players to migrate between different clans and alliances, these clans and alliances do not have much space for institutional innovation and no need to comply with democratic principles of governance. Also, the underlying business model with significant subscription fees leading to a degree of income-based segregation and the tendency of the games to only appeal to a non-representative subset of population limit their relevance.

Nonetheless, these limitations are perhaps removable. Could democratic innovation be gamified? Could people’s hands be untied to try various institutional designs and move between groups governed by different rules, triggering competition and selection of the most workable mechanisms? Perhaps it would be possible to adapt some of the already developed digital worlds to serve as a space for democratic innovation where people are given freedom and incentives to explore various institutional frameworks to decide for themselves which of them best facilitate their cooperation. The hope is that such

exploration will generate data relevant for assessing the promise of the real-world political reforms much more efficiently than what has been heretofore possible.

Admittedly, my proposal is just a highly preliminary sketch. However, it might hint at a passable way towards ultimately reconstituting democratic decision-making to better cope with anthropic risks in the digital age.

Conclusion

Nowadays, anthropic risks are significant and steadily growing. However, their increasing urgency does not prove democracy obsolete or inadequate. Rather than falling once again for the pipedream of benevolent despotism that just ‘cuts to the chase and does what is necessary,’ we should seek ways to manage anthropic risks with the people’s participation and consent.

True, liberal democracy may not be resilient enough against the current tides of political irrationality to prove itself up to the task, especially as the epistemic democratization driven by the digitalization of the political discourse upsets its habitual ways of information dissemination. Nonetheless, the proper answer is not abandoning democracy but seeking a different democratic institutional template for political decision-making.

Given the rigidity of the status quo institutions, facilitating an exploration of democracy’s design space becomes of the essence. I propose creating democracy’s digital playground, where different designs could compete with one another. Institutional exploration using digital worlds like those we already know from massively multiplayer online games remains an understudied option that just might work.

Unfortunately, the clock is ticking fast. Time is a luxury that humanity does not currently have in abundance. In the realm of collective choice, further reliance on the mechanisms that performed so poorly in the face of climate change and the COVID-19 pandemic appears exceedingly adventurous. We need to get our act together. Otherwise, anthropic risks are bound to produce disastrous outcomes.

References

- Brennan, J. (2016). *Against Democracy*. Princeton: Princeton University Press.
- Christiano, T. (2011). An Instrumental Argument for a Human Right to Democracy: An Instrumental Argument for a Human Right to Democracy. *Philosophy & Public Affairs*, 39(2), 142–176. <https://doi.org/10.1111/j.1088-4963.2011.01204.x>

- Croce, M. & Piazza, T. (2021). Consuming Fake News: Can We Do Any Better? *Social Epistemology*, 1–10. <https://doi.org/10.1080/02691728.2021.1949643>
- Epley, N. & Gilovich, T. (2016). The Mechanics of Motivated Reasoning. *Journal of Economic Perspectives*, 30(3), 133–140. <https://doi.org/10.1257/jep.30.3.133>
- Fishkin, J., Siu, A., Diamond, L. & Bradburn, N. (2021). Is Deliberation an Antidote to Extreme Partisan Polarization? Reflections on “America in One Room.” *American Political Science Review*, 1–18. <https://doi.org/10.1017/S0003055421000642>
- Gelfert, A. (2018). Fake News: A Definition. *Informal Logic*, 38(1), 84–117. <https://doi.org/10.22329/il.v38i1.5068>
- Goodin, R. E. & Spiekermann, K. (2018). *An epistemic theory of democracy*. Oxford, United Kingdom: Oxford University Press.
- Gurri, M. (2018). *The revolt of the public and the crisis of authority in the new millennium*. San Francisco: Stripe Press.
- Hayek, F. A. (1945). The Use of Knowledge in Society. *The American Economic Review*, 35(4), 519–530.
- Hertwig, R. & Grüne-Yanoff, T. (2017). Nudging and Boosting: Steering or Empowering Good Decisions. *Perspectives on Psychological Science*, 12(6), 973–986. <https://doi.org/10.1177/1745691617702496>
- Kahan, D. M. (2017). Misinformation and identity-protective cognition. *Fake News in Context*, 51–57.
- Kahneman, D. (2003). Maps of Bounded Rationality: Psychology for Behavioral Economics. *The American Economic Review*, 93(5), 1449–1475. <https://doi.org/10.2307/3132137>
- Landemore, H. (2012). *Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many*. Princeton: Princeton University Press.
- Landemore, H. (2020). *Open democracy: Reinventing popular rule for the twenty-first century*. Princeton: Princeton University Press.

- Mercier, H. (2020). *Not born yesterday: The science of who we trust and what we believe*. Princeton NJ: Princeton University Press. <https://doi.org/10.1515/9780691198842>
- Oliver, A. (2013). From Nudging to Budging: Using Behavioural Economics to Inform Public Sector Policy. *Journal of Social Policy*, 42(4), 685–700. <https://doi.org/10.1017/S0047279413000299>
- Ord, T. (2020). *The precipice: Existential risk and the future of humanity*. New York: Hachette Books.
- Posner, E. A. & Weyl, E. G. (2018). *Radical markets: Uprooting capitalism and democracy for a just society*. Princeton: Princeton University Press.
- Repucci, S. (2021). Freedom in the World 2020. A Leaderless Struggle for Democracy. Retrieved May 19, 2021, from Freedom House website: <https://freedomhouse.org/report/freedom-world/2020/leaderless-struggle-democracy>
- Simler, K. & Hanson, R. (2018). *The Elephant in the Brain: Hidden Motives in Everyday Life*. New York: Oxford University Press.
- Sismondo, S. (2017). Post-truth? *Social Studies of Science*, 47(1), 3–6. <https://doi.org/10.1177/0306312717692076>
- Špecián, P. (2019). The Precarious Case of the True Preferences. *Society*, 56(3), 267–272. <https://doi.org/10.1007/s12115-019-00363-8>
- Špecián, P. (2022a). *Behavioral political economy and democratic theory: Fortifying democracy for the digital age*. New York, NY: Routledge.
- Špecián, P. (2022b). Epistemology and the Pandemic: Lessons from an Epistemic Crisis. *Social Epistemology*, 36(2), 167–179. <https://doi.org/10.1080/02691728.2021.2009931>
- Sunstein, C. R. (2017). *#Republic: Divided Democracy in the Age of Social Media*. Princeton ; Oxford: Princeton University Press.
- Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Anchor Books.
- Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New York: Yale University Press.

Funding

The participation in the panel “Risks in the Anthropocene” within the scope of the *Green Marble 2022 - International Meeting on Anthropocene Studies and Ecocriticism* and the work on this paper have been supported by the Czech Science Foundation (GACR) under Grant 20-06678S.