WORKING PAPER SERIES • Vol. 6 • 2024:10-17 •

STUDIES ON

CLIMATE ETHICS

AND FUTURE GENERATIONS

Working paper series 2024:10–17

Editors: Julia Mosquera & Olle Torpman

Studies on Climate Ethics
and Future Generations
Vol. 6

Studies on Climate Ethics and Future Generations Vol. 6

Editors: Julia Mosquera & Olle Torpman

Institute for Futures Studies
Working Papers 2024:10-17
Stockholm 2024

The Institute for Futures Studies is an independent research foundation financed by contributions from the Swedish Government and through external research grants. The institute conducts interdisciplinary research on future issues and acts as a forum for a public debate on the future through publications, seminars and conferences.

© The authors and the Institute for Futures Studies 2024

Cover: Matilda Svensson/Tove Salomonsson
Cover image: Yosh Ginsu/Unsplash
Print: Elanders, Stockholm, 2024
Distribution: The Institute for Futures Studies, 2024

Contents

Preface

The Climate Ethics and Future Generations programme is now completed. It was led by PI Gustaf Arrhenius and co-PIs Krister Bykvist and Göran Duus-Otterström from 2018 to 2024, hosted by the Institute for Futures Studies (IFFS) in Stockholm and financed by Riksbankens Jubileumsfond. The primary objective was to integrate and develop the most important insights regarding climate ethics from different subject areas and methodologies. It has merged normative research from philosophy, economics, and political science with empirical research from economics, sociology, and demography. A research environment was created in which climate ethics researchers from all over the world worked together across academic disciplines. The programme's core team comprised 33 researchers but also involved a great number of other researchers involved in the preprints, workshops and conferences arranged by the programme, in total six preprints (available at iffs.se and climateethics.se), 32 workshops and four scientific conferences.The programme established IFFS as a world leading hub for climate ethics research.

The program's research has been disseminated through 231 scientific articles, 70 chapters in anthologies and handbooks, and over 300 scientific presentations. Eleven scientific books have been published or accepted for publication, including Population Ethics, Moral Uncertainty, and What We Owe Future People, by OUP. The researchers have been actively participating in outreach activities, resulting in over 200 interviews, panel discussions, lectures, and op-eds. In a notable collaboration with the art world, the researchers participated in the creation of the 8-meter high performative sculpture Tipping Point.

Many new and exciting research questions emerged from the programme, as evidenced by the fact that it has generated nineteen on-going major spin-off projects. These are new research projects that have received external funding to further explore the questions and research environment created by the programme. Some examples are: The effect of climate change on non-human animals; how to manage catastrophic climate risk; severe empirical uncertainty; how should individuals, groups, and states coordinate their actions to mitigate climate change; ethical questions concerning the positive discount rate used in integrated assessment models; the feasibility and efficacy of so-called climate clubs.

The programme had three broad themes: Foundational questions in population ethics, which concerns how we should evaluate future scenarios in which the number of people, their welfare, and their identities may vary; Climate justice, which concerns the just distribution of the burdens and benefits of climate change and climate policy, both intra- and intergenerationally; and From theory to practice,

which concerns how to apply normative theories to the circumstances of climate change, in light of both normative uncertainty and practical constraints. For more information about the program, visit climateethics.se.

The three themes are duly represented in this sixth and final volume of the programme's preprint series, consisting of eight papers in total. The papers are presented in alphabetic order of the author names.

The volume's opening paper, "Productive Justice in the 'Post-Work Future'", Caleb Althorpe and Elizabeth Finneron-Burns explore the distribution of work benefits and burdens in a future where technological unemployment might be widespread. The authors argue that while non-work benefits (like income) can be met elsewhere, social contribution remains unique to work and central to justice. They contend that, given technology's limits in replacing care work, equitable distribution of care burdens is essential. Egalitarian principles, therefore, require a balance between work relief and shared responsibilities in care tasks.

The second paper, "Degrees of Incommensurability and the Sequence Argument", by Gustaf Arrhenius and H. Orri Stefánsson, addresses Parfit's Sequence Argument against the Repugnant Conclusion, focusing on new notions of incommensurability proposed by Hájek and Rabinowicz. The proposal avoids the Repugnant Conclusion only by allowing extreme weight on inequality. This approach leads to a dilemma, where, under some views, a population with universally better welfare can still be deemed no better than a less advantaged one, posing challenges to intuitive and ethical population rankings.

Next, Katharina Berndt Rasmussen explores, in her paper "Discrimination and Future Generations", whether current generations' actions—like resource depletion—constitute discrimination against future generations. After defining discrimination and its moral implications, she argues that future generations' temporal detachment limits this claim, especially given the non-identity problem. While intergenerational inequities exist, they lack the grounds needed for discrimination classification, as future people do not form a socially comparable community with present-day people.

In the fourth paper, "Escaping the Impossibility Theorems in Population Ethics", Krister Bykvist investigates the impossibility theorems in population ethics, which challenge the development of fair policies addressing climate impacts. He proposes a flexible approach, treating the satisfaction of ethical conditions as degrees rather than absolutes. Using the Kemeny measure, he suggests that some population theories may be preferable based on their distance from paradoxical outcomes, offering a nuanced way to approach conflicting ethical demands in policy.

The fifth paper of the volume, "Generationally Parochial Geoengineering", by Stephen M. Gardiner and Catriona McKinnon, argues that geoengineering initia-

tives, like sulfate injections, risk being short-sightedly biased towards the immediate generation's interests. The authors examine ethical concerns about neglecting future generations' wellbeing in current geoengineering discourse. They call for heightened ethical scrutiny in geoengineering decisions, advocating for more inclusive considerations to prevent generational injustice and broader environmental risks.

Next, Clare Heyward and Edward Page discuss how some climate policies, despite addressing primary environmental issues, inadvertently create "secondary injustices," causing further harm to certain communities. In their paper, "Rectifying Secondary Climatic Injustices", Heyward and Page argue for compensatory measures for those affected and examine factors like policymakers' awareness of alternative approaches. The goal is an equitable distribution of climate policy burdens, particularly when initial policies inadvertently worsen existing inequalities.

The seventh paper, "Sufficiency and the Distribution of Burdens", by Robert Huseby, critiques sufficientarianism's tolerance of inequality above a minimum welfare threshold, especially concerning climate-related burdens. Huseby explores revisions to sufficientarian views that would require fair burden-sharing beyond mere sufficiency. His proposed adjustments aim to maintain the core sufficientarian principles while addressing concerns about unjust burden allocations that might disadvantage those just above the sufficiency threshold.

In the volume's final paper, "Benefiting at the Expense of Climate Change", Edward Page examines the ethical obligations of individuals and entities profiting from activities that exacerbate climate change. He differentiates between unjust and wrongful enrichment and argues that while unjust enrichment may not justify legal recuperation, wrongful enrichment serves as a basis for moral duties to rectify gains that come at the planet's expense. This concept offers a normative foundation for accountability in climate justice beyond straightforward legal frameworks.

We are pleased to be able to share this work from the Climate Ethics and Future Generations project. As with previous volumes, the authors of these papers would greatly appreciate any comments, questions, and objections that you wish to share with them. Contact information is found at the front page of each paper. We would also like to thank Erika Karlsson for assisting with formatting the papers in this volume.

Julia Mosquera & Olle Torpman
Editors

Caleb Althorpe[1] & Elizabeth Finneron-Burns[2]

Productive Justice in the 'Post-Work Future'[3]

Justice in production is concerned with ensuring the benefits and burdens of work are distributed in a way reflective of persons' status as moral equals. While a variety of accounts of productive justice have been offered, insufficient attention has been paid to the distribution of work's benefits and burdens in the future. In this paper, after granting for the sake of argument forecasts of widespread future technological unemployment, we consider the implications this has for egalitarian requirements of productive justice. We argue that in relation to all the benefits affiliated with work, other than undertaking social contribution, the technological replacement of work is unproblematic as these benefits could in principle be attained elsewhere. But because social contribution uniquely corresponds to work (when work is understood as more than a paid job), the normative assessment technological unemployment will turn on the value theories of justice give to contributive activity. We then argue that despite technological replacement being plainly beneficial insofar as it relieves persons from the burdens of work, such as dangerous work or drudgery, because the nature of care work makes it less susceptible to technological replacement, egalitarian concern will require the burdens of care work to be shared equally between individuals.

[1] Caleb Althorpe, IRC Government of Ireland Postdoctoral Fellow, Philosophy Department, Trinity College Dublin (althorpc@tcd.ie)

[2] Elizabeth Finneron-Burns, Assistant Professor, Department of Political Science, University of Western Ontario (efinnero@uwo.ca)

[3] This paper is forthcoming at the Journal of Applied Philosophy. We are grateful to them for permission to preprint it here.

1. Introduction

It is no less true today than in the past that technological advancement changes the nature and availability of work. From the perspective of workers, some of these contemporary changes might be for the better, while others might be for the worse. For instance, increased computing power has led to it being more feasible for individuals to undertake remote work with its affiliated flexibility. On the other hand, the use of algorithms in organizations often disconnects workers from important details of the work process or can lead to organizational changes making work more precarious (like gig work). Because of the position of work in the economies of today, most people will spend more time working than doing anything else in their life. This means any changes to work's organization, nature, and its availability in society will have significant impacts on how individuals' lives fare overall. The changes brought on the work process by technological advancement, then, are relevant to theories of social justice.

In this paper we focus on the normative issues surrounding one such (predicted) change: the potential of technological advancement to lead to widespread automation and unemployment in the future. This prediction about the ability of technology to bring about a future in which the majority of work (paid and unpaid) is automated, we will call the technological assumption.

Several studies have suggested that approximately half the work currently undertaken in advanced economies could be fully automated in the near-term,[4] while others predict that it is a real possibility that in a matter of decades automation will be so widespread that most individuals will no longer be able to work for money.[5] Of course, historical predictions about the effects of technological advancement on the amount of work in society have been notoriously wrong,[6] and one does not need to look too far to find contemporary skepticism towards the technological assumption.[7] In reply, what advocates of the technological assumption claim is that the displacement potential of technology is qualitatively different than it was in the past, for two reasons. First, new technologies like artificial intelligence (AI) and machine learning models can replace not just complex physical tasks but complex mental ones too. Second, the rate of change in digital technologies increases exponentially.[8]

---

[4] Frey and Osborne, "The Future of Employment"; Manyika et al., *A Future that Works.*

[5] Chace, *The Economic Singularity*; Ford, *Rise of the Robots.*

[6] For some examples, see Autor, "Why are there still so many jobs?", 3–4.

[7] Denning, "The 'Jobless Future' is a Myth"; Atkinson and Wu, "False Alarmism"; Spencer, "Fear and Hope"; Benanav, "A World Without Work".

[8] Ernst, Merola, and Samaan, "Economics of Artificial Intelligence," 3; Danaher, Automation and Utopia, 30–48.

Some examples might help give us a sense of this displacement potential: trading algorithms and AI decision-support tools have replaced much of the work in the finance sector; machine learning outputs often give better medical diagnoses than human doctors; smart robots are beginning to carry out surgeries independently from human doctors; AI now provides reliable legal advice on the likelihood of winning court cases; many news articles are now completely written by algorithms; and many companies have fully automated their customer service through AI chat bots.[9] Even persons in creative occupations, like artists and poets, are vulnerable to replacement by technological automation, given the recent (and forecasted future) developments in deep learning models (as seen in tools like DALL-E and ChatGPT). We will refer to a world in which most work is automated as a 'post-work future', while recognising that not all work can be automated (more on that later).

This paper's exploration of the normative issues related to a post-work future is not motivated by accepting the technological assumption as inevitably true or guaranteed, but by acknowledging that the prediction is a non-zero possibility; we are granting the technological assumption for the sake of argument. If the future were a world where technological advancement has meant most people no longer work, would this be a good or bad thing from the standpoint of justice? Some think such a world would be a kind of utopia,[10] but is that right? It is important to get clear on this question, as it will be normatively action-guiding in the present. Since the technological assumption is predicted to materialise in the medium term, a significant number of those affected by it have not even been born yet. If a post-work future would be an unqualified good thing, then perhaps efforts should be made to maximize technological development in order to benefit future people. Or if such a future would lead to the occlusion of certain benefits for future people, then perhaps it gives us reason to put the brakes on the technological advancement, or to at least explore alternative means through which these benefits could be attained outside of work.

While the topic of a post-work future (or at least job displacement) receives significant attention in public discourse and from social scientists and technology ethicists, it receives comparatively little focus from political theorists and political philosophers, and we hope to begin to rectify that here. The prospect of a post-work future is a topic of concern to economic justice because it relates to the distribution of work's benefits and burdens and how society's productive activity is organized and carried out. It is also of concern to intergenerational justice because it is a question of what social institutions the current generation either leave or bring about for

[9] Susskind and Susskind, The Future of the Professions, 46-100; Ford, "The Rise of Robots," 35–38; Danaher, Automation and Utopia, 7–20.

[10] Bastani, Fully Automated Luxury Communism; Danaher, Automation and Utopia.

future generations, and normative assessment of a post-work future will depend on whether the benefits and burdens of work for individuals in the present will remain benefits and burdens for individuals in the future.

The paper proceeds as follows. We begin in Section II by defining what we should take 'work' and a 'post-work future' to mean. In Sections III and IV we then examine five things that are often taken to be benefits of work and argue that four of them are not inherent in work, but rather contingent on it, so could still be realised in the post-work future. The fifth benefit we take to be inherent in work but argue that there are reasons to think it might no longer be normatively significant in a post-work world, so it too will not necessarily be a reason to prevent the post-work world from materialising. In Section V, starting from the fact that even in the post-work future some work would remain (viz., affective care work), we argue that this remaining work creates concerns central to productive justice. If the post-work world is to be an egalitarian one, then technological displacement must be accompanied by positive efforts to ensure the remaining labour is distributed fairly.

2. Work and the 'Post-Work World'

What do we mean by 'work,' and hence what do we mean by a post-work world brought about by technological advancement?

By work we mean more than a paid job, and we follow several accounts in understanding work as activity that meets others' needs insofar as it generates goods or services that are useful or necessary for others being able to carry out their (reasonable) plan of life.[11] Seeing work in this way, as social contribution that is useful to others, does a good job capturing the sort of activities commonly regarded as 'work'. The account captures market-facing work (such as the paid work undertaken within employment relations and by independent market actors) due to the information function of the price mechanism – if the activity were not useful to others (or at least expected to be useful to others), then nobody would pay for it.[12] The account also

[11] Van Parijs, Real Freedom for All, 138; Tilly and Tilly, Work Under Capitalism, 22; Cholbi, "The Duty to Work", 1122; Geuss, A Philosopher Looks at Work, 18.

Resultantly, throughout the paper we characterize work as being 'useful' and as meeting needs interchangeably. Equating meeting needs with useful activity in this way makes it broader than an account of basic needs. You obviously do not need the ice cream you buy in order to survive, but it is useful to you because it helps you carry out whatever aims and plans you have chosen to prioritize (you might use it to relax after a long day's work or need to take it to a friend's dinner party, and so on...). We think defining work in terms of needs in this way is attractive because positive social contributions through work surely capture more than just those things persons strictly need to survive (the ice cream maker is making a social contribution). But at the same time, by remaining objective, it stops work simply becoming activity that meets any and all subjective wants, no matter how unreasonable. We thank an external reviewer for asking us to elaborate on this point.

[12] Carens, Equality, Moral Incentives, and the Market, 195; Van Parijs, Real Freedom for All, 138; Brown,

captures non-market-facing work activities, such as unpaid domestic and care work, and volunteer work, because each of these activities produces goods and services that are necessary and useful to others. These latter activities – which are disproportionately undertaken by members of disadvantaged groups – are still work despite their going unpaid (because society both racializes certain work and devalues what is regarded as 'women's work'[13] or because the market fails to produce public goods, or whatever). It is social contribution that explains why we want to call domestic labour and caring 'work', but not reading or going for a jog. This account of work as activity related to what other persons need, also captures how several philosophers treat work as an inherently necessary activity, and that this is what separates work from leisure given the latter has value only for the person or people doing it.[14]

With this understanding of work in mind, we are characterizing the post-work society predicted by the technological assumption as not just a society where robots and AI have come in and replaced paid jobs. Rather, we are understanding the post-work society as a society where technology has displaced the majority of both paid *and* unpaid work. It is a society that no longer requires most people to engage in *any* activities that are useful or necessary to others (with some exceptions we will detail later). It is understandable that most of the public concerns about technological displacement relate to paid jobs, given most persons' means to a livelihood is the income they receive through work. However, because there are also nonpecuniary benefits to work, it is the scenario where technology has displaced the complete set of work activities that needs normative assessment. A post-work society of this kind is clearly not right around the corner. Resultantly, our focus is on what justice might say about the prospect of technological displacement in the medium to long-term, and our paper is silent on normative issues surrounding the impact of technology on work processes in the present and near-term future.[15]

3. Four Benefits of Work that Won't be Missed

We identify five distinct benefits that political theorists and philosophers commonly associate with the work activity: income, self-development and excellence, community, meaningfulness, and social contribution. We do not take these five benefits as an exhaustive and complete list of work's benefits,[16] but they are the benefits most

"The Meaning of Markets," 232.

[13] Daniels, "Invisible Work," 404–405.

[14] Rose, Free Time, 37; Clark, "Good Work," 62-63; Cholbi, "Philosophical Approaches to Work and Labor".

[15] Vredenburgh, "The Right to Explanation"; Bankins and Formosa, "The Ethical Implications of Artificial Intelligence".

[16] For example, see Arneson, "Meaningful Work and Market Socialism," 528-529; Arneson, "Is Work

often put forward as normatively relevant or of concern to theories of social justice. In characterizing the benefits (and burdens) of work we aim to remain neutral in relation to three major factors that differentiate alternative accounts of justice: the appropriate metric of justice (resources, opportunities, capabilities, welfare, etc.), the distributive rule of justice (equality, priority, sufficiency, etc.), and the relationship between justice and the good (perfectionism, nonperfectionism). We do this to delineate what is normatively relevant about the technological assumption to accounts of economic justice generally.

(i) Income

Perhaps the most immediately obvious benefit of (much) work is that it serves as a means to an income. Work has exchange value insofar as individuals can sell their labour to an employer, or their work products or services to buyers in the market. Money received through work is clearly relevant to justice-motivated concerns with individuals' material prospects and income inequality. Rawls' difference principle, to just take one example, measures how persons fare in terms of the income they receive through work.[17] Pecuniary benefits from work are of course a prototypical case of an extrinsic benefit – the benefit is only what results from work and has nothing to do with features of the work process itself.[18]

(ii) Excellence

But there are also benefits to work that are internal to the work process. The first of these is how undertaking work is connected to individuals' self-development and the attainment of excellence. It is often through work that persons can best accomplish tasks that depend upon the deployment of their developed skills and talents (be they physical, mental, or emotional). A factor that makes work a natural place for self-development is the limits to what can be achieved in a single life (a person can't be all at once a top-tier athlete, a master writer, and a talented therapist). There is then something of a social division of labour between the particular skills and achievements individuals choose or have the capacity to develop.[19]

Self-development and the attainment of excellence in work is taken as a justice-relevant benefit for a variety of reasons. The most familiar one might be accounts that give priority to self-realization and skill deployment in work as part of a view of human flourishing, be this in terms of an Aristotelian account of human capacities,[20]

Special?" 1132.

[17] Rawls, *A Theory of Justice*, 78, 96-98; Rawls, *Justice as Fairness,* 63.

[18] See Cholbi, "Philosophical Approaches to Work and Labor".

[19] Rawls, *A Theory of Justice,* 23–525.

[20] Clark, "Good Work".

or a Marxian account of persons being connected to their species-being through skilled work.[21] But accounts need not be perfectionist to see self-development and excellence in work as justice-relevant, given the "internal resources" of intelligence and virtuosity cultivated through skilled work can be regarded in the interests of persons generally since they are useful in other realms of life.[22] Furthermore, self-development might be taken as a benefit of work because it is connected to increases in individual welfare and enjoyment, and acts as a major motivating factor for individuals choosing some types of work over others.[23] Regardless of the reasons that self-development is taken to be justice-relevant, the institutional upshot for accounts of economic justice is to prioritize work processes that have a degree of complexity and which give scope for agency. These work processes are antithetical to work that is drudgery, such as when there is a detailed horizontal division of labour resulting from work being organized according to principles of scientific management.

(iii) Community

The work process is also a common way for individuals to attain the good of community. This is especially so for individuals who work not as independent market actors but as employees of organizations, where interactions with colleagues and shared involvement in a collective project that is valued can foster relations of sociability and cooperation.[24] Examples might be mechanics in an auto shop each deploying their own expertise to fix a tricky issue, doctors and psychologists working together to help a sick patient, a collective of artists expressing beauty each in their own way, and even philosophers working together to advance knowledge. Insofar as workers value the work-related end of their activity (the mechanics value maintaining cars for its own sake, and so on) then workplaces can be forms of community. Just like self-realization and excellence, community in and through work can be valued for a variety of reasons. Community at work can be taken as important because it provides workers with a context in which their skills can receive recognition and appraisal, and hence be taken as worthwhile.[25] But it can also be seen as valuable in more 'political' terms, where relations of community and solidarity in workplaces are valued because they foster a sense of the common good, the latter which forms part of the democratic virtues that maintain political stability.[26] The concern with

[21] Attfield, "Work and the Human Essence"; Elster, "Self-Realization in Work and Politics".

[22] Arnold, "The Difference Principle at Work".

[23] Gheaus and Herzog, "The Goods of Work," 75 and the references there.

[24] Estlund, *Working Together*, 3-7; Gheaus and Herzog, "The Goods of Work," 76.

[25] Doppelt, "Rawls' System of Justice", 275–276.

[26] O'Neill, "Three Rawlsian Routes Towards Economic Democracy", 42–48.

community in work most commonly takes aim at hierarchical relations. Anca Gheaus and Lisa Herzog for instance, while acknowledging some forms of organizational hierarchy are surely legitimate, argue that from the standpoint of community, workplace democracy and worker cooperatives are the ideal form of workplace organization.[27]

(iv) Meaningfulness

Meaningfulness or 'meaningful work' is another commonly identified benefit of work. While some writers regard meaningful work just as work that enables the other benefits of work to be attained[28], meaningfulness through work is often taken as a distinct kind of benefit. Having confidence that one's work is significant, purposeful and extends 'beyond the self' in some way is one common descriptor of meaningful work[29], while other writers understand meaningful work as work that gives the worker scope to exercise autonomy and agency.[30] Many accounts take meaningful work to be important as part of a larger claim about the ethical significance of individuals having a secure sense of meaning in life more generally[31], while other writers characterize meaningful work in less philosophically demanding terms and see it as valuable merely out of its connection to persons' political status and their sense of self-worth.[32] For these latter writers, the 'meaningful' in meaningful work is understood not in terms of fundamental meaningfulness or meaning in life, but instead only in terms of what might make work as a distinct activity meaningful.[33] Regardless of the exact way meaningful work is characterized, institutional implications for accounts of justice that value meaningful work include guarantees of complex and interesting work, as well as work that gives workers a democratic say in managerial decisions.[34]

Why will these four benefits not be missed in a post-work world? If income, self-development and excellence, community, and meaningfulness (and the values with

[27] Gheaus and Herzog, "The Goods of Work," 77-78; Schwarzenbach, "Rawls and Ownership", 149–150, 162–163.

[28] E.g., Gheaus and Herzog, "The Goods of Work", 71.

[29] Fried and Ferris, "The Validity of the Job Characteristics Model"; Grant, "Relational Job Design"; Lips-Wiersma and Morris, "Discriminating Between 'Meaningful Work' and the Management of Meaning".

[30] Schwartz, "Meaningful Work"; Roessler, "Meaningful Work: Arguments from Autonomy".

[31] Yeoman, "Conceptualising Meaningful Work as a Fundamental Human Need"; Veltman, *Meaningful Work*; Tyssedal, "Good Work".

[32] Moriarty, "Rawls, Self-Respect, and the Opportunity for Meaningful Work"; Althorpe, "Meaningful Work, Nonperfectionism and Reciprocity".

[33] See Althorpe, "What is Meaningful Work?", 587–588.

[34] Esheté, "Contractarianism and the Scope of Justice", 43; Schwartz, "Meaningful Work," 639–642; Hasan, "Rawls on Meaningful Work and Freedom," 481-482; Breen, "Meaningful Work and Freedom", 59–61.

which they are affiliated) are all relevant to concerns of social justice, won't a world where these benefits are no longer attainable through work be a bad thing? To see why not, we need to recognize that these benefits are only *contingently* connected to work – there is no inherent connection. This is obviously true of income but it is also true of the nonpecuniary benefits that are internal to the work process, and is something of which several accounts of economic justice are aware—these benefits are only benefits of *work* because we spend so much of our time working. Each of them can, at least in principle, be realised outside of work. As put by Gheaus and Herzog, "[w]e would have less, if any, reason, to be concerned with the distribution of the nonmonetary goods of work if we were to reform employment such that people spent much less time in paid work and had more time flexibility".[35] Clearly, the post-work future envisaged by the technological assumption is one such reform, and so long as technological advancement occurs alongside some kind of policy providing individuals a guaranteed revenue stream (one common example being a universal basic income funded by an automation tax[36]), the benefits outlined above would still be available to persons living in a post-work world.

First of all, as a basic income shows, income can obviously be provided in ways other than compensation for work. But what such a post-work world also does is open up opportunities to engage in and derive benefits from non-work pursuits. Without work sapping much of persons' energy and effort, they could devote their (much increased) leisure time to personal projects, hobbies, and interests, all which could involve significant skill development and the deployment of talents. While such talents would no longer be as closely tied to social necessity, such activities would still enable the values affiliated with self-development and excellence to be realized. Just to take one example, even if in the post-work world sport entertainment was provided by robots, what is relevant to accounts that give value to excellence and self-development it is the fact that people will still be able to maximize their potential and develop their capacities as an athlete. Similarly, while community with colleagues will no longer be an option in a post-work world, community will be possible with friends, or with fellow hobby enthusiasts, or in religious organizations, and so on, given these are also avenues for persons to engage in shared activity relating to collectively valued ends. Finally, at least when meaningful work is understood in terms of meaning in life generally, this will not be unique to work because work is not the only sphere through which a person's activity can extend 'beyond the self' in the relevant sense – this can just as easily occur in things like democratic and political participation, religious beliefs, art, literature, or even

[35] Gheaus and Herzog, "The Goods of Work," 80.

[36] Bruun and Duka, "Artificial Intelligence, Jobs and the Future of Work".

philosophical reflection. If anything, we might think that meaningfulness in these realms of life is going to be more important to individuals than any meaningfulness derived from work, given non-work activities are often more closely tied to people's personal conceptions of the good or beliefs about fundamental value than any work activity can be (similar comments might apply to the prospects of community in non-work activities and the subsequent recognition and appraisal received). It is true however that if the 'meaningful' in meaningful work is understood only in terms of what might make work as a distinct activity meaningful (and not in terms of meaning in life), then it won't be available in a post-work future. But as far as we can ascertain, when meaningful work is understood in this way then any benefit it is taken to have relies on it extending the work process 'beyond the self' by giving workers an opportunity to use their developed skills to positively contribute to others.[37] This means that concerns about the availability of the benefit of meaningful work so understood fold into concerns about opportunities for social contribution, which we consider in the next section.

Therefore, at least in relation to these four benefits of work, the technological displacement of work will not be a problem for future people (so long as it occurs alongside the provision of something like a universal basic income). Indeed, we might even have reason to think that opening up the range of activities through which these benefits could be attained is something justice requires, given that having them available only through the work activity (when an alternative possibility is available) would be privileging one kind of conception of the good and way of life over others.[38] This is perhaps especially so for theories of justice that have a nonperfectionist bent, but even if the benefits were taken as valuable in perfectionist terms it seems at odds with such an approach to limit the ways through which the good or human flourishing can be acquired. Maximizing the opportunities future people have for excellence, say, requires opportunities for excellence are available across a wide range of activities, not just work.[39]

[37] Just as one example, take for instance the way Elizabeth Anderson characterizes meaningful work as: "work that affords a means for a person to exercise their agency and skill in the course of helping other people" ("The Struggle for Meaningful Work," 75). See also Hasan, "Rawls on Meaningful Work and Freedom," 503–504.

[38] Birnbaum, "Should Surfers be Ostracized?", 400-403; Weeks, *The Problem with Work,* 97–103; Jenkins, "Everybody's Gotta Do Something"; Beverinotti, "Beyond Work: Life, Death, and Reproduction and the Postwork Society", 264–266.

[39] E.g., see Wall, "Perfectionist Justice and Rawlsian Legitimacy," 423–424.

4. One Benefit of Work that Might be Missed

(v) Social Contribution

The final benefit of work that is often mentioned is social contribution. Different accounts of labour locate the value of social contribution in different places. Some accounts regard social contribution as objectively valuable, either for perfectionist or nonperfectionist reasons. Examples of the former are accounts that prioritize the value of unalienated labour and how this is tied to work that not only 'completes' the worker, but which also 'completes' and is appreciated by its beneficiary[40], and those that give positive value to pro-sociality.[41] An example of the latter might be when social contribution is taken as valuable because it relates to persons' political status as members of society characterized as a system of social cooperation.[42] Other accounts, meanwhile, take social contribution as valuable in terms of its relation to individuals' subjective attitudes, where the emphasis is put on the idea that it is only through work individuals are able meet their desire to contribute to and help others.[43]

Regardless of which account you accept, unlike the previous four benefits of work the benefits affiliated with social contribution are not merely contingent on work but are inextricably linked to the work activity. This naturally results from what we argued above was the most convincing description of work – activity that is useful or necessary for others to carry out their plan of life. Therefore, while just like with the benefits considered in the previous section, a post-work society will bring about a scenario where this benefit is no longer attainable through work (because there isn't much work), since social contribution is inherent to the work process and not merely contingent to it, this means that it will not be available through other kinds of activities like the other benefits will be.

Before considering the normative implications of this, we will first respond to the rejoinder that even if social contribution is inherent to work, in a post-work world there will still be sufficient opportunity to undertake activities that are useful to others, and so the benefits affiliated with social contribution can be retained. One way to characterize the idea could be to say that while individuals might not be able to contribute to others through working, they will be able to contribute to others by playing games (in Bernard Suits' sense, where games are "the voluntary attempt to overcome unnecessary obstacles"[44]). Indeed, several writers think it likely the play-

[40] Brudney, "Two Marxian Themes"; Kandiyali, "The Importance of Others",

[41] Tyssedal, "Good Work".

[42] Althorpe, "Meaningful Work, Nonperfectionism, and Reciprocity".

[43] Gheaus and Herzog, "The Goods of Work," 75.

[44] Suits, *The Grasshopper,* 41.

ing of games would become a dominant activity in a post-work world,[45] and the thought might go that because such games will bring pleasure and provide an end to ourselves and others, they will contribute in that way. For example, we might invite a depressed friend out for a round of golf to cheer him up, or we might even play with them the game of 'housebuilding' or 'taxi driving.'

But this line of thought misunderstands the nature of social contribution tied to the work activity. Work as a form of social contribution is not just about doing all the things that can be useful to our immediate social circles (friends and family with whom we would play games), but about doing the things that are necessarily useful to people with whom we are unassociated.[46] This is true even for domestic and care work because raising a child (for example) is useful not just to the child, but to society at large. Playing golf with your friend and raising a child might both be useful to others, but only the latter is a form of social reproduction and contribution (society depends upon the rearing in a way it doesn't depend upon the golf game between friends). Given the scenario of technological displacement under consideration here, the game of 'housebuilding' is no more necessary from a social point of view than playing golf. If someone really needed a roof over their heads, then they would get the robots to make them one.

Given then that the post-work world will deprive us of the benefit of social contribution, does that mean that any accounts of social justice that give normative weight to the act of social contribution have reason to object to the technological displacement of work? It appears that they might, and that this derives from an obligation to prevent future people being deprived of a justice-relevant benefit. Such an outcome would after all be based on the same normative considerations (e.g., the value of unalienated labour, or the way social contribution is tied up with self-worth) that underpin the way such accounts criticize how contemporary relations of work fall short of what justice requires.

But while our aim in this paper is not to interrogate the merits of this or that account of economic justice, we do think the technological assumption might give us reason to be skeptical of using the premises on which these accounts base the normative significance of social contribution to criticize the prospects of a post-work future. After all, the attractiveness of these claims about the benefit of social contribution must at some point fall back on claims about the inherent interdependence between persons (as otherwise the value given to social contribution seems arbitrary). Marxian accounts, for instance, characterize the importance of unalienated labour that completes others and situates the worker closely to social contribu-

[45] Suits, *The Grasshopper,* ch. 15; Black, *The Abolition of Work and Other Essays*; Danaher, "In Defense of the Post-Work Future".

[46] Althorpe, "What is Meaningful Work?".

tion in terms of persons producing in a "human manner" that "objectif[ies] the human essence".[47] And while connecting social contribution to the human essence might be plausible in the here and now (we are making no judgement about that), it seems such a connection would be significantly undermined in a future world where robots are able to do the majority of the productive work. And if interdependence (through undertaking activities useful to others) is no longer inherent to the human essence, Marxian-style arguments that prioritize the benefit of social contribution appear to lose much of their normative thrust. While some writers bite the bullet here by claiming a future world where robots do the vast majority of the work would no longer be a human society,[48] this is just begging the question. These accounts have a burden of proof to show why our human essence couldn't be defined by some other feature.

Similar comments apply to accounts that value social contribution in nonperfectionist terms by connecting social contribution to the characterization of society as a system of cooperation. Such a characterization explains why social contribution is normatively significant insofar as it is connected to persons having a secure sense of self-worth as participating members of society, or persons satisfying their desires to meet others' needs and be useful to others. At least in the present, it may well be reasonable to care about contributing socially since we (accurately) see ourselves as part of a reciprocal system where everyone is required to do their part through work. But in a post-work world where machines will be doing the majority of work, the idea that social contribution will continue to be constitutive of society as a system of cooperation will surely be undermined in the same way as any account of human essence based on social contribution.[49] And if social contribution is no longer tied to the features of political society, then there seems no reason to think it ought to be tied to persons' sense of worth or self-respect as members of society, or be connected to desires to contribute that would matter to an account of justice that is focused on the provision of all-purpose means. The ideals we have currently, as producers, or of society as a system of cooperation, might be reasonable and provide justification for individuals in the here and now, but this might not be the case for people in the long term, post-work future.

It is helpful here, we think, to note how many writers criticize the normative weight given to work as problematically ideological. Common forms of this criticism are that beliefs about work's value are just an unhelpful historical carry-over from

[47] Marx, "On James Mill", 132. See also Brudney, "Two Marxian Themes"; Kandiyali, "The Importance of Others".

[48] E.g., Deranty, "Post-Work Society as an Oxymoron", 426–427.

[49] Of course, society may still cooperate for other beneficial reasons, for example, by all obeying the law in order to maintain safety and security.

pre-industrial society, or result from an updated secularized version of the Protestant worth ethic where persons continue to uncritically prioritize and internalize duties towards work and beliefs about its importance. And the argument goes that insofar as processes such as these explain the continuing importance given to work and social contribution, then such beliefs are unjustified or at best misplaced, and we need to move beyond them.[50] As Richard Arneson puts it in discussing the way contemporary society ties social esteem and status to work, this is just a cultural belief that could be changed, and "perhaps an egalitarian norm ought to reject this way of thinking".[51] What these writers emphasize is that we can surely define ourselves as humans and derive our purpose and self-worth in the spheres of life that exist outside of work and social contribution.[52]

We do not raise this line of argument because we think all contemporary valuation of work is necessarily ideological, but because we think it is hard to deny that this criticism has a lot of bite when applied to the scenario of a post-work future. What gets counted as a justice-relevant benefit ought to be sensitive to changing social conditions. And the potential 'transcending' of interdependence through technological development that the post-work future promises is a such a significant change that we need to be very careful that any objection to its development is not in effect imposing a set of values that might be appropriate in one time and place onto individuals who will (or could) live in a very different world.

To sum up the discussion thus far, four benefits of work were found not to be inherent in work itself, but are rather a result of the sheer amount of time individuals currently spend in work. In a post-work future, these benefits would be realisable through other activities undertaken in significantly increased discretionary time. However, because the benefits affiliated with social contribution are inherent to work itself, these could not be generally realised in a post-work future where the majority of work is done by machines. While this might initially appear to be one reason to object to the prospects of a post-work future, we argued that there are good reasons to think the overcoming of the inherent interdependencies the technological assumption claims can be brought about, would result in social contribution becoming significantly less valuable in a post-work world. The consequence of this analysis is that there is likely no reason, from a benefits-of-work point of view, to object to the technological assumption materialising. In the next section, however, we argue that given there is one kind of work (affective care work) that is likely to remain in the post-work future, and because the changes brought about by the tech-

[50] E.g., Russell, "In Praise of Idleness"; Frayne, *Refusal of Work*. But see generally Muirhead, *Just Work,* 95-113; Deranty, "Post-Work Society as an Oxymoron"fab, 105–111.

[51] Arneson, "Is Work Special?", 1133.

[52] Weeks, *The Problem with Work,* 230–233.

nological assumption are unlikely by themselves to undermine norms and expectations around who ought to do this work, there are egalitarian reasons to ensure that the technological displacement of work in the future is accompanied by positive efforts to ensure the labour that remains is distributed fairly.

5. Equality in a Post-Work Future

In order to consider whether the benefits of work commonly identified can be used to object to a post-work future generally, we have so far considered the effects/value of work on people in a noncomparative sense, assuming that the effects of the technological assumption will apply equally. However, a full assessment of the technological assumption from the standpoint of economic justice will need to also take into account that members of different social categories are differently situated to the institution of work. Indeed, the nature of these social categories and how they relate to others are often intimately linked to work. For example, some argue that the reason care work often goes unpaid is because historically it has been done primarily by women and has therefore been undervalued by patriarchal societies.[53] Another example is the theory of racial capitalism that claims social categories of race play a functional role in justifying the unequal consequences of capitalist systems and operate in ways that maintain their stability.[54] The final question we want to interrogate, therefore, is whether the post-work future is likely to disrupt, rely on, or reproduce social orderings that are unjust.

One potential positive of a post-work future is that automation can relieve individuals from undertaking the burdens associated with certain kinds of work – burdens which currently fall disproportionately more on some groups in society over others. For example, what is currently considered dangerous and 'dirty' work is often the easiest to automate. Fishing, mining, working on oil rigs, and construction are just a few examples of jobs in which workers are regularly injured and/or killed. Garbage collection, sewage treatment, and some medical professions like personal support workers are examples of 'dirty' jobs in which workers are exposed to unpleasant smells, sights, or others' bodily fluids and functions. These are obvious burdens to the work process, burdens which are not merely the absence of the goods outlined earlier. If these dangerous and dirty jobs are automated, not only would it be a good thing that people no longer needed to perform dangerous or dirty work, but also, due to the demographics of who tends currently to be subject to the burdens affiliated with these roles, this would have positive effects on redressing an

[53] Daniels, "Invisible Work".

[54] Robinson, *Black Marxism, Revised and Updated*; Bright et al., "On the Stability of Racial Capitalism".

existing inequality. In the United States for example, Black and Hispanic workers are much more likely to work in high-risk occupations than are white workers. The unfortunate consequence of this is that Black and Hispanic workers are 39% and 27% more likely, respectively, to be injured at work than a white worker.[55] By reducing racial disparities like this one, in this regard the post-work future would undoubtedly be a good thing from the standpoint of racial equality.

However, as we alluded to earlier in this paper, we think it is a mistake to regard the post-work future as a world where *all* work will be eliminated, and we argue that there is one class of work that will inevitably remain – affective care work (such as childcare, elder care and the like). This means questions about this work's fair distribution will very much still be live in a post-work future. By *affective* care work, we are utilizing the distinction between 'functional' and 'affective' care.[56] Functional care refers primarily to meeting people's physical needs—cleaning, feeding, moving people, for example. Affective care refers to meeting people's emotional needs—lending a sympathetic ear, helping them with problems, loving them.[57] Robots are likely, in the future, to be able to perform most forms of functional care. After all, we already have things like self-emptying robot vacuums and mops, dishwashers, and self-cleaning ovens, so it is no stretch to imagine machines taking over the tasks such as diaper changes and meal preparation for children, and there being self-driving cars ushering them off to their myriad of extracurricular activities. When it comes to affective care work however, the potential of technological displacement is far less certain. This is because essential components of good affecttive care include conscious attentiveness, deep empathy and respect, and reciprocity[58] and there still appears to be a significant gap when it comes to the ability of machines to replicate emotional states such as these (in contrast to their ability to replicate physical and mental tasks). As AI philosopher Robert Sparrow has put it, "robots cannot provide genuine care because they cannot experience the emotions that are integral to the provision of such care"[59].

---

[55] Seabury, Terp and Boden, "Racial and Ethnic Differences in the Frequency of Workplace Injuries and Prevalence of Work-Related Disability".

[56] Coghlan, "Robots and the Possibility of Humanistic Care".

[57] While there is some overlap, affective care is not the same as emotional labour as the latter refers to the specific way certain jobs require employees to manage and regulate their expressions and personae in customer interactions and encapsulates a broader set of work than 'care work' (think flight attendants, hotel concierges, and so on. See Hochschild, *The Managed Heart*. Given the emotional states at issue in some types of emotional labour are less demanding than those in affective care work (the hotel company only wants you to feel welcomed, not understood), the prospects of robots providing it is more plausible. Hence, we are not arguing that all forms of emotional labour cannot be automated, only that affective care work cannot be automated (without losing part of what makes it a social contribution).

[58] E.g., see Tronto, Moral Boundaries.

[59] Sparrow, "Robots in aged care," 449. See also Vallor, "Carebots and Caregivers"; Sharkey and Sharkey,

Take for instance the work of raising a child. For this to effectively meet the child's needs (but also the needs of others given the role of child rearing in social reproduction), it needs to rely on loving them, spending time with them, and caring *about* them, not just taking care *of* them (their physical needs). Or if we turn to elder care, this is about much more than merely feeding, dressing, and cleaning those who are no longer independent, but about listening to individuals' stories, chatting with them, keeping them company, and letting them know that someone cares about them and empathizes with them as they age. In both these cases, what proper care requires is the affective attention that reflects that those cared for are owed respect, consideration, and dignity, and which shows that they are valued as ends in themselves.

Therefore, while robots might be able to meet the physical needs affiliated with the functional tasks commonly constitutive of care work, given the lack of human intersubjectivity they will be unable to meet the emotional needs affiliated with affective care. This means if robots fully replaced human care workers, then this would significantly reduce the extent the activity is a social contribution. Given what we argued earlier about this being what makes something work in the first place (Section II), such a result would not be *displacing* human work with machine work, it would be *removing* the work altogether, given the needs of others are no longer being met.[60]

Furthermore, it is important to consider that functional care and affective care, particularly of children and the elderly, cannot be easily separated, meaning that even if we had robots capable of performing functional care, we could not allow them to do so without sacrificing a significant amount of affective care. To see why this is so, let's return to the examples from the previous paragraph. When a parent takes care of a young baby, the vast majority of day-to-day care is functional—changing diapers, bathing them, and breast or bottle feeding them. However, when a parent does these tasks, they also engage with the baby. They talk to her, have skin-to-skin contact with her, make eye contact, etc. When a parent drives their child to baseball practice, they are not just providing transportation, but they are also talking with their child, engaging with them, showing them that they care, and bonding. Changing diapers and driving children provide both functional and affective care for the child because they benefit her physically *and emotionally* and further the bond between parent and child. Studies show that children's time spent with their parents

"The Rights and Wrongs of Robot Care"; Stokes and Palmer, "Artificial Intelligence and Robotics in Nursing"; Coghlan, "Robots and the Possibility of Humanistic Care". Cf. DeFalco, "Towards a Theory of Posthuman Care".

[60] None of this is meant to deny that there could be a place for machines in the effective provision of affective care, just that there could never be a total displacement of human labour without a cost to social contribution.

positively correlates with better educational outcomes, less contact with the criminal justice system, less substance abuse, and higher self-esteem.[61] A robot completing then even functional care or supervision of children, let alone emotional tasks (affective care), would not be meeting needs as effectively as it would be unlikely to have the same outcomes for the child, parent, or child-parent bond. The upshot then is that even if the vast majority of functional care were automated (the technological possibility of which is plausible), there would still be a set of emotional needs that only affective care undertaken by humans could meet. The post-work future then is not one where all human labour has been replaced.

What are the implications of all this for concerns of productive justice? Although we have been careful to refer to care workers with the neutral 'they,' this work currently primarily falls to women. The vast majority of unpaid care work is currently undertaken by women, and most *paid* care workers are women, in particular, immigrant women and women of colour.[62] The initial question to ask then is what effect the radical transformation of labour brought about by the technological assumption might have on this gendered (and racialized) division of labour.[63]

One might be optimistic and think that the post-work world is likely to remove gender inequality in virtue of relieving women of many burdens of care. This might occur through two mechanisms. First, at the moment, when men are offered paid parental leave, they take it. Perhaps this suggests that men, when given the opportunity to care (for children in this case), choose to do so, and in the post-work future, when much more of their time is freed up, men will choose to engage in much more care. Second, since affective care work is likely to be the only remaining opportunity for obtaining the benefit of social contribution through work, it is possible that men will develop more of an interest in performing it. We, however, are more pessimistic. While we don't want to deny that the changes brought about by the technological assumption might result in some improvement to the current unequal distribution of care, we think it very unlikely such mechanisms will make concerns with the distribution of care in the post-work world irrelevant.

Regarding first the claim—that men, once they have the time to do so, will choose to spend their time caring—unfortunately, studies do not bear this out. It is true that,

[61] Wikle and Cullen, "The Developmental Course of Parental Time Investments in Children".

[62] Lum, Sladek and Ying, "Ontario Personal Support Workers in Home and Community Care".

[63] Due to space constraints, in what follows we have chosen to focus on the gendered aspect of the unequal division of care work. The way racial oppression manifests in care work is, of course, complex (e.g., see Bhandary, "Caring for Whom? Racial Practices of Care and Liberal Constructivism"). But given one significant reason racial minorities and migrant workers are more likely to undertake care work is because of a lack of meaningful economic alternatives, then a universal basic income in the post-work future will likely go some way towards the reduction of that inequality, and perhaps be more effective than in relation to the gendered division of care insofar as the latter is tied to economic inequality to a lesser degree.

when paid parental leave is offered (or sometimes mandated), men take it up. However, rather than spending the leave time caring for their babies, they tend to use it to upskill, take on extra work for additional income, explore new business ideas, and/or look for new career opportunities.[64] Furthermore, studies have shown that when women are the sole income-earners and the men are stay-at-home fathers, men still do less childcare (19 hours per week) than their working female partners (21 hours per week).[65] So even when men are relieved of their paid work, they still do not do as much care as their (paid) working female partners. This strongly suggests that lack of time is not the impediment to men's participation in care work, and we think makes it reasonable to expect that any additional time afforded by the post-work future is unlikely to significantly change, on its own, the gendered division of care.

Let's now consider the possibility that men will be more likely to take up care work because it offers the last remaining option to obtain the benefits associated with social contribution. First of all, we have already raised doubts about the importance of social contribution in the post-work future (Section IV) and argued that it is certainly possible that the bases of persons' self-respect, or their 'essence' as humans, could be found in realms of life outside of work. Applied to the point here, while men may no longer be able to base their self-worth on being a breadwinner, perhaps they will be able base it on how, just to take one example, they perform in the games they now play with their friends. But even if social contribution continues to remain a benefit, we don't think this is enough to warrant thinking that inequalities in who does the care work will be overcome. This is because there is no reason to think the benefits of social contribution will necessarily be taken as special or more weighty than other benefits. 'Sure,' our imaginary individual might think, 'I might get some benefit from undertaking my fair share of affective care work, but think of all the more freedom I will have to do what I want if I leave this socially necessary labour to others.' Again, we don't want to suggest the change in social circumstances brought about by the post-work future will result in no progress towards gender equality. But given how deeply gendered norms regarding care work have been entrenched in social institutions historically and in the present, should we really expect the increase in free time brought about by technological development to be enough on its own to overcome this?

The takeaway is this: as we have described it, the post-work world will be one in which almost all work, save some care work, is automated. And given we have raised some doubts about the ability of this post-work world to overcome, by itself, norms

[64] Tharp and Parks-Stamm, "Gender Differences in the Intended Use of Parental Leave".

[65] In dual income earning families, women do 23 hours of childcare per week compared to 12 hours for men. Baxter, "Stay-at-home-dads (Facts Sheet).

about gender roles and the division of affective care, this amounts to a world in which women will disproportionately work (at care) whilst men will disproportionately enjoy post-work leisurely lives. This would be unjust. And this is so even if through this unequal division of the remaining labour women will have on average greater access to the benefits affiliated with social contribution, given these benefits only result from socially-imposed norms and expectations.[66]

This suggests that an essential consideration to ensure a post-work world is consistent with productive justice will be positive efforts to bring about the egalitarian division of affective care work. One way to bring this about could be mandatory participation in affective care, along the lines of Elizabeth Brake's Care Corps or Cécile Fabre's civilian service.[67] With systems such as these, each individual would be required to do their fair share of socially necessary care work (depending on one's personal circumstances and the community's needs, this might be within the family or for strangers). This would then ensure that the non-automated care work that people need for their lives to go well is done equitably. If, after each person performs their fair share of care work, some choose to perform more (perhaps in line with women's 'natural' desire to care), that would be supererogatory and not a problem from the standpoint of gender equality.

If we are wrong about men's preferences on average in the post-work future and men and women are equally likely to want to do the care work that remains, then this policy would not be coercive or freedom-limiting at all. However, if there were divergent preferences, then it would involve a degree of coercion, and resultantly could seemingly be regarded as inconsistent with other values often thought constitutive of justice in production (like free choice of occupation). But the response here is to fall back on the fact that such coercion is only necessary because certain people have unreasonable preferences—viz., the preference men have to freeride on the care work performed by others in order to carry out their own lives however they see

[66] An anonymous reviewer puts the following objection to us: the very fact that care work is distributed in a gendered way does not mean that it is necessarily unjust. Even in a society devoid of gender discrimination and social norms concerning different kinds of work, it's possible that women might still be more drawn to care work than men. Even if that is true (though we doubt this), the care work involved is a form of socially necessary labour so is still heteronomous to some extent, at least compared to leisurely pursuits. Therefore, even though there can be positive goods associated with it, the necessity of the work and heteronomy suggests that it is still unjust have one class of people disproportionately undertaking the work for the benefit of the other, especially when the latter do no work at all and can do as they please. The latter would essentially be freeriding on the work done by the former.

[67] Brake, "Fair Care: Elder Care and Distributive Justice"; Fabre, *Whose Body Is it Anyway?*.

One concern with this approach might be that people needing care may end up cared for by people who are not very good at or interested in caring for others. This is a legitimate concern that merits further consideration, particularly with respect to how it should be balanced against the unjustness of the gendered division of labour.

fit. Not only would this policy have the result of an equitable division of the care work that remains in society, but it might also contribute to undermining existing gendered norms surrounding care work in the first place. Children would be guaranteed to grow up being cared for by both men and women, girls and boys would be taught to care in school and in the home and would grow up with the expectation that they will do so in equal amounts in adulthood. These social factors, alongside the explicit state-sanctioned message that care work is performed by women *and men,* would likely, over time, significantly reduce the need for coercion in the first place.

6. Conclusion

The aim of this paper has been to begin to normatively assess the value of a 'post-work future' and outline what considerations of justice arise from the technological displacement of work. We have argued that the post-work future should not be rejected simply because it would result in the loss of the benefits of work because most of those benefits are only contingent on work and can be realised in other ways, given the significant increase in discretionary time that will characterise the post-work world. We also argued that although the benefit of social contribution could not be realised outside of work, there are reasons to be skeptical that it would continue to be a meaningful benefit at all in a world of automation. Although the loss of benefits of work are not reasons to prevent the technological assumption from materialising, it is also important to consider how different social groups are situated to the institution of work differently. What we have focused on, is that given affective care work is likely to resist automation, and because the technological changes bringing about a post-work future are unlikely to undermine gendered norms and expectations about this work, then when we are thinking about the design of institutions in the post-work world, concerns of productive justice and gender equality are inseparable.

Acknowledgements

We would like to thank two anonymous reviewers and audiences at the "Future of Work" workshop at the Centre for Ethics and Philosophy in Practice (ZEPP) at LMU Munich and the "Future of Work and Income" workshop at the Centre for Ethics, Philosophy, and Public Affairs at the University of St. Andrew's for very helpful comments that significantly improved the paper. We also appreciate funding from the Social Sciences and Humanities Research Council of Canada, Riksbankens Jubileumsfond, and the Irish Research Council.

References

Althorpe, Caleb. 2022. "Meaningful Work, Nonperfectionism, and Reciprocity." *Critical Review of International Social and Political Philosophy*. Online First. https://doi.org/10.1080/13698230.2022.2137753

Althorpe, Caleb. 2023. "What Is Meaningful Work?" *Social Theory and Practice* 49(4): 579–604.

Anderson, Elizabeth. 2023. "The Struggle for Meaningful Work," *Dissent* 70(3): 69–75.

Arneson, Richard. 1987. "Meaningful Work and Market Socialism." *Ethics* 97(3): 517–545.

Arneson, Richard. 1990. "Is Work Special? Justice and the Distribution of Employment." *The American Political Science Review* 84(4): 1127–1147.

Arnold, Samuel. 2012. "The Difference Principle at Work." *Journal of Political Philosophy* 20(1): 94–118.

Atkinson, Robert D. and John Wu. 2017. "False Alarmism: Technological Disruption and the U.S. Labor Market, 1850-2015." *ITIF@Work Series*. Online. Available at: https://www2.itif.org/2017-false-alarmism-technological-disruption.pdf

Attfield, Robin. 1984. "Work and the Human Essence." *Journal of Applied Philosophy* 1(1): 141–150.

Autor, David H. 2015. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." *Journal of Economic Perspectives* 29(3): 3–30.

Bankins, Sarah and Paul Formosa. 2023. "The Ethical Implications of Artificial Intelligence (AI) For Meaningful Work." *Journal of Business Ethics*. Online First. https://doi.org/10.1007/s10551-023-05339-7

Bastani, Aaron. 2019. *Fully Automated Luxury Communism: A Manifesto*. London: Verso.

Baxter, Jennifer. 2017. Stay-at-home dads (Facts Sheet). Melbourne: Australian Institute of Family Studies.

Benanav, Aaron. 2020. "A World Without Work?" *Dissent* 67(4): 45–52.

Beverinotti, Matías. 2021. "Beyond Work: Life, Death, and Reproduction and the Postwork Society." *Discourse* 43(2): 264–285.

Bhandary, Asha Leena. 2022. "Caring for Whom? Racial Practices of Care and Liberal Constructivism." *Philosophies* 7(4): 78–91.

Birnbaum, Simon. 2011. "Should Surfers be Ostracized? Basic Income, Liberal Neutrality, and the Work Ethos." *Politics, Philosophy & Economics* 10(4): 396–419.

Black, Bob. 1986. *The Abolition of Work and Other Essays*. Port Townshend: Loompanics Unlimited.

Brake, Elizabeth. 2017. "Fair Care: Elder Care and Distributive Justice" *Politics, Philosophy & Economics* 16(2).

Breen, Keith. 2019. "Meaningful Work and Freedom: Self-Realization, Autonomy, and Non-Denomination in Work," in *The Oxford Handbook of Meaningful Work*, eds., Ruth Yeoman, Catherine Bailey, Adrian Madden, and Marc Thompson. New York: Oxford University Press, 51–72.

Bright, Liam Kofi, Nathan Gabriel, Cailin O'Connor, and Olúfẹmi O. Táíwò. Forthcoming. "On the Stability of Racial Capitalism." *Ergo: An Open Access Journal* of Philosophy. https://doi.org/10.31235/osf.io/syk4m

Brown, Brookes. 2022. "The Meaning of Markets," in *The Routledge Handbook of Philosophy, Politics, and Economics*, ed., C. M. Melenovsky. New York: Routledge, 227–238.

Brudney, Daniel. 2018. "Two Marxian Themes: The Alienation of Labour and the Linkage Thesis." In *Reassessing Marx's Social and Political Philosophy*, ed. Jan Kandiyali. New York: Routledge, 211–238.

Bruun, Edvard P.G. and Alban Duka. 2018. "Artificial Intelligence, Jobs and the Future of Work: Racing with the Machines." *Basic Income Studies* 13(2): 1–15.

Carens, Joseph H. 1981. *Equality, Moral Incentives, and the Market: An Essay in Utopian Politico-Economic Theory*. Chicago: University of Chicago Press.

Chace, Calum. 2016. *The Economic Singularity: Artificial Intelligence and the Death of Capitalism*. Three Cs.

Cholbi, Michael. 2018. "The Duty to Work." *Ethical Theory and Moral Practice* 21(5): 1119–1133.

Cholbi, Michael. 2023. "Philosophical Approaches to Work and Labor." *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), eds., Edward N. Zalta and Uri Nodelman. https://plato.stanford.edu/entries/work-labor/

Clark, Samuel. 2017. "Good Work." *Journal of Applied Philosophy* 34(1): 61–73.

Coghlan, Simon. 2022. "Robots and the Possibility of Humanistic Care," *International Journal of Social Robotics* 14(10): 2095–2108.

Danaher, John. 2019a. *Automation and Utopia: Human Flourishing in a World Without Work*. Cambridge: Harvard University Press.

Danaher, John. 2019b. "In Defense of the Post-Work Future: Withdrawal and the Lucid Life." In "The Future of Work, Technology, and Basic Income," eds., Michael Cholbi and Michael Weber. New York: Routledge, 113–130.

Daniels, Arlene Kaplan. 1987. "Invisible Work." *Social Problems* 34(5): 403–415.

DeFalco, Amelia. 2020. "Towards a Theory of Posthuman Care: Real Humans and Caring Robots" *Body & Society* 26(3): 31–60.

Denning, Steve. 2015. "The 'Jobless Future' is a Myth." *Forbes*. Online. Available at: https://www.forbes.com/sites/stevedenning/2015/06/04/the-robots-are-not-coming/

Deranty, Jean-Philippe. 2015. "Historical Objections to the Centrality of Work," *Constellations* 22(1): 105–121.

Deranty, Jean-Philippe. 2022. "Post-Work Society as an Oxymoron: Why We Cannot, and Should Not, Wish Work Away." *European Journal of Social Theory* 25(3): 422–439.

Doppelt, Gerald. 1981. "Rawls' System of Justice: A Critique from the Left." *Noûs* 15(3): 259–307.

Elster, Jon. 1986. "Self-Realization in Work and Politics: The Marxist Conception of the Good Life." *Social Philosophy & Policy* 3(2): 97–126.

Ernst, Ekkehardt, Rossana Merola, and Daniel Samaan. 2019. "Economics of Artificial Intelligence: Implications for the Future of Work." *IZA Journal of Labor Policy* 9(4): 1–35.

Esheté, Andreas. 1974. "Contractarianism and the Scope of Justice." *Ethics* 85(1): 38–49.

Estlund, Cynthia. 2003. *Working Together: How Workplace Bonds Strengthen a Diverse Democracy*. Oxford: Oxford University Press.

Fabre, Cécile. 2006. *Whose Body is it Anyway?* Oxford: Oxford University Press.

Ford, Martin. 2016. *Rise of the Robots: Technology and the Threat of a Jobless Future*. New York: Basic Books.

Ford, Martin. 2018. "The Rise of Robots: Impact on Unemployment and

Inequality." In *Confronting Dystopia: The New Technological Revolution and the Future of Work*, ed., Eva Paus. Ithaca: Cornell University Press, 27–45.

Frayne, David. 2015. *Refusal of Work: The Theory and Practice of Resistance to Work*. London: Zed Books.

Frey, Carl Benedikt and Michael A. Osborne. 2017. "The Future of Employment: How Susceptible are Jobs to Computerisation?" *Technological Forecasting & Social Change* 114: 254–280.

Fried, Yitzhak and Gerald R. Ferris. 1987. "The Validity of the Job Characteristics Model: A Review and Meta-Analysis." *Personnel Psychology* 40(2): 287–322.

Gheaus, Anca and Lisa Herzog. 2016. "The Goods of Work (Other Than Money!)" *Journal of Social Philosophy* 47(1): 70–89.

Geuss, Raymond. 2021. A Philosopher Looks at Work. Cambridge: Cambridge University Press.

Grant, Adam M. 2007. "Relational Job Design and the Motivation to Make a Prosocial Difference." Academy of Management Review 32(2): 393–417.

Hasan, Rafeeq. 2015. "Rawls on Meaningful Work and Freedom." Social Theory & Practice 41(3): 477–504.

Hochschild, Arlie Russell. 2003. The Managed Heart: Commercialization of Human Feeling, 20th Anniversary Edition. Berkeley: University of California Press.

Jenkins, David. 2020. "'Everybody's Gotta Do Something': Neutrality and Work." Critical Review of International Social and Political Philosophy 23(7): 831–852.

Kandiyali, Jan. 2020. "The Importance of Others: Marx on Unalienated Production." Ethics 130(4): 555–587.

Lafargue, Paul, 1907. "The Right to Be Lazy," in The Right to Be Lazy and Other Studies, trans., C. H. Kerr. Chicago: Charles H. Kerr & Co., 3–62.

Lips-Wiersma, Marjolein and Lani Morris. 2009. "Discriminating Between 'Meaningful Work' and the 'Management of Meaning.'" Journal of Business Ethics 88(3): 491–511.

Lum, Janet, Jennifer Sladek and Alvin Ying. 2010. "Ontario Personal Support Workers in Home and Community Care", Canadian Research Network for Care in the Community, Toronto Metropolitan University.

Manyika, James et al. 2017. A Future that Works: Automation, Employment, and Productivity. McKinsey Global Institute Report. Online. Available at: https://www.mckinsey.com/~/media/mckinsey/featured insights/Digital Disruption/Harnessing automation for a future that works/MGI-A-future-that-works-Executive-summary.ashx

Marx, Karl. 2000e. "On James Mill." In Karl Marx: Selected Writings (2nd edition), ed., David McLellan. New York: Oxford University Press, 124–133.

Moriarty, Jeffrey. 2009. "Rawls, Self-Respect, and the Opportunity for Meaningful Work." Social Theory and Practice 35(3): 441–459.

Muirhead, Russell. 2004. Just Work. Cambridge: Harvard University Press.

O'Neill, Martin. 2008. "Three Rawlsian Routes Towards Economic Democracy." Revue de Philosophie Économique 8(2): 29-–55.

Rawls, John. 1971. A Theory of Justice (Original Edition). Cambridge: Belknap Press.

Rawls, John. 2001. Justice as Fairness: A Restatement. Cambridge: Belknap Press.

Robinson, Cedric J. 2020. Black Marxism, Revised and Updated Third Edition: The Making of the Black Radical Tradition. Chapel Hill: The University of North Carolina Press.

Roessler, Beate. 2012. "Meaningful Work: Arguments from Autonomy." Journal of Political Philosophy 20(1): 71–93.

Rose, Julie L. 2016. Free Time. Princeton: Princeton University Press.

Russell, Bertrand. 1983. "In Praise of Idleness." In Why Work? Arguments for the Leisure Society, ed., Vernon Richards. London: Freedom Press.

Schwartz, Adina. 1982. "Meaningful Work." Ethics 92(4): 634–646.

Schwarzenbach, Sibyl A. 1987. "Rawls and Ownership: The Forgotten Category of Reproductive Labor." Canadian Journal of Philosophy Supplementary Volume 13: 139–167.

Seabury, Seth, Sophie Terp and Leslie Boden. 2017. "Racial And Ethnic Differences In The Frequency Of Workplace Injuries And Prevalence Of Work-Related Disability." Health Affairs 36(2).

Sharkey, Noel and Amanda Sharkey. 2012. "The Rights and Wrongs of Robot Care," in Robot Ethics, eds., Patrick Lin, Keith Abney, and George A. Bekey. Cambridge: MIT Press, 267–282.

Sparrow, Robert. 2016. “Robots in aged care: a dystopian future?” AI & Society 31: 445–454.

Spencer, David A. 2018. “Fear and Hope in an Age of Mass Automation: Debating the Future of Work.” New Technology, Work, and Employment 33(1): 1–12.

Stokes, Felicia and Amitabha Palmer. 2020. “Artificial Intelligence and Robotics in Nursing: Ethics of Caring as a Guide to Dividing Tasks Between AI and Humans,” Nursing Philosophy 21(4): 1–9.

Suits, Bernard. 1978. The Grasshopper: Games, Life, and Utopia (1st ed.). Toronto: University of Toronto Press.

Susskind, Richard and Daniel Susskind. 2015. The Future of the Professions: How Technology Will Transform the Work of Human Experts. Oxford: Oxford University Press.

Tharp, Derek T. and Elizabeth Parks-Stamm. 2021. “Gender Differences in the Intended Use of Parental Leave: Implications for Human Capital Development” Journal of Family and Economic Issues 42: 47–60.

Tilly, Chris and Charles Tilly. 1998. *Work Under Capitalism*. New York: Routledge.

Tronto, Joan. 1993. Moral Boundaries: A Political Argument for an Ethic of Care. New York: Routledge.

Tyssedal, Jens Jørund. 2022. “Work is Meaningful if There are Good Reasons to do it: A Revisionary Conceptual Analysis of ‘Meaningful Work.’” Journal of Business Ethics. Online First. https://doi.org/10.1007/s10551-022-05205-y

Tyssedal, Jens Jørund. 2023. “Good work: The importance of caring about making a social contribution.” Politics, Philosophy & Economics Online First. https://doi.org/10.1177/1470594X221148204

Vallor, Shannon. 2011. “Carebots and Caregivers: Sustaining the Ethical Ideal of Care in the Twenty-First Century,” Philosophy & Technology 24(3): 251–268.

Van Parijs, Philippe. 1995. Real Freedom for All: What (if Anything) Can Justify Capitalism? Oxford: Clarendon Press.

Veltman, Andrea. 2016. Meaningful Work. New York: Oxford University Press.

Vredenburgh, Kate. 2022. “The Right to Explanation.” The Journal of Political Philosophy 30(2): 209–229.

Wall, Steven. 2014. “Perfectionist Justice and Rawlsian Legitimacy.” In A Companion to Rawls, eds., Jon Mandle and David A. Reidy. Chichester: Wiley Blackwell, 413-429.

Weeks, Kathi. 2011. The Problem with Work: Feminism, Marxism, Antiwork Politics, and Postwork Imaginaries. Durham: Duke University Press.

Wikle, Jocelyn and Clara Cullen. 2023. “The Developmental Course of Parental Time Investments in Children from Infancy to Late Adolescence” Social Sciences 12(2).

Yeoman, Ruth. 2014. “Conceptualising Meaningful Work as a Fundamental Human Need.” Journal of Business Ethics 125(2): 235–251.

Gustaf Arrhenius & H. Orri Stefánsson[1]

Degrees of Incommensurability and the Sequence Argument

Parfit (2016) responded to the Sequence Argument for the Repugnant Conclusion by introducing imprecise equality. However, Parfit's notion of imprecise equality lacked structure. Hájek and Rabinowicz (2022) improved on Parfit's proposal in this regard, by introducing a notion of degrees of incommensurability. Although Hájek and Rabinowicz's proposal is a step forward, and may help solve many paradoxes, it can only avoid the Repugnant Conclusion at great theoretical cost. First, there is a sequential argument for the Repugnant Conclusion that uses weaker and intuitively more compelling assumptions than the Sequence Argument, and which Hájek and Rabinowicz's proposal only undermines, in a principled way, by allowing for seemingly implausible weight to be put on the disvalue of inequality. Second, if Hájek and Rabinowicz do put such seemingly implausible weight on the disvalue of inequality, then they will have to accept that a population A is not worse than another population B even though everyone in B is better off than anyone in A.

[1] Institute for Futures Studies and Stockhom University

1. Introduction

Here's a simple and general formulation of Derek Parfit's infamous "Repugnant conclusion":

> *The Repugnant Conclusion*: For any population consisting of people with very high positive welfare, there is a better population in which everyone has a very low positive welfare, other things being equal.[2]

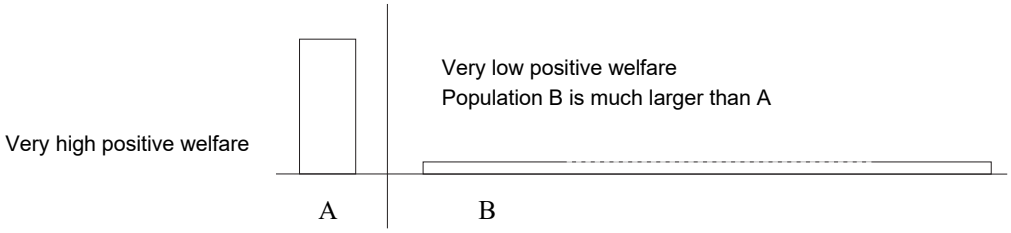


Diagram 1

In diagram 1, the width of each block represents the number of people whereas the height represents their lifetime welfare. Dashes indicate that the block in question should be much wider than shown, that is, the population size is much larger than shown.

These populations could consist of all the past, present and future lives (a possible world), or all the present and future lives, or all the lives during some shorter time span in the future such as the next generation, or all the lives that are causally affected by, or consequences of a certain action or series of actions, and so forth.[3]

All the lives in the diagram have positive welfare, or, as we also could put it, all the people have lives worth living. The A-people have very high welfare whereas the B-people have very low positive welfare.[4] The reason for this could be that in the B-

[2] For Parfit's original formulation, see Parfit (1984), p. 388.Our formulation is more general than his. For early sources of the Repugnant Conclusion, see Arrhenius (2000b), (2016), (forthcoming).

[3] More exactly, a population is a finite set of lives in a possible world. A, B, C, ..., A_1, A_2, ..., A_n, $A \cup B$, and so on, denote populations of finite size. We shall adopt the convention that populations represented by different letters, or the same letter but different indexes, are pairwise disjoint. For example, $A \cap B = A_1 \cap A_2 = A' \cap B' = \varnothing$. We shall assume that for any natural number *n* and any welfare level **X**, there is a possible population of *n* people with welfare **X** (for a discussion of this *No-Limit Assumption*, see Arrhenius (2000b) ch. 3, (forthcoming)).

[4] For a discussion and definition of positive, negative, and neutral welfare, see Arrhenius (2000b), (forthcoming) ch. 2 and 9 (for a short summary, see Arrhenius (2016)). Cf. Broome (1999), (2004), Bykvist (2007), p. 101, and Parfit (1984), pp. 357–358 and appendix G. Notice also that we actually don't need an analysis of a neutral welfare in the present context but rather just a criterion, and the criterion

lives there are, to paraphrase Parfit, only enough ecstasies to just outweigh the agonies, or that the good things in those lives are of uniformly poor quality, e.g., eating potatoes and listening to Muzak.[5] However, since there are many more people in B, the total sum of welfare in B is greater than in A. Hence, a theory like Total Utilitarianism, according to which we should maximize the welfare in the world, ranks B as better than A – an instance of the Repugnant Conclusion.[6]

Notice that the Repugnant Conclusion is not just a problem for total utilitarians or those committed to welfarism – the view that welfare is the only value that matters from the moral point of view – since the *ceteris paribus* clause in the formulation implies that the compared populations are equal in all possibly axiologically relevant respects apart from individual welfare levels. Hence, other values and considerations are not decisive for the value comparison of populations A and B. Thus, the Repugnant Conclusion is a problem for all moral theories according to which welfare matters at least when all other things are equal, which arguably is a minimal adequacy condition for any moral theory.[7]

As the name indicates, Parfit found the Repugnant Conclusion very counterintuitive and most philosophers seem to agree. However, there is a well-known and tempting argument for the Repugnant Conclusion, which Parfit called the "Continuum" Argument. That is an unfortunate misnomer, since the argument does not in fact require a continuum. Therefore, we shall instead refer to it as the "Sequence Argument". In section II we explain the Sequence Argument in more detail, but in short, the argument starts with a population like A, where everyone has very high positive welfare, and then introduces a sequence of populations, where each population is much bigger but offers slightly lower individual welfare than the previous population in the sequence. One might hold that for any two consecutive populations in this sort of sequence, the latter, if sufficiently large, is better than the former much smaller one, since the reduction in individual welfare is so small. But then, since "better than" is a transitive relation, we sooner or later get the Repugnant Conclusion, that is, we find that a population like B, in Diagram 1, must be better than population A in Diagram 1.

Parfit (2016) responded to the Sequence Argument by suggesting that adjacent populations are actually "imprecisely equally good". In section II we briefly explain Parfit's response, but the important observation about imprecise equality is that it

---

can vary with different theories of welfare.

[5] See Parfit (1984), p. 388 and Parfit (1986), p. 148. For a discussion of different interpretations of the Repugnant Conclusion see Arrhenius (2000b), (forthcoming) and Parfit (1984), (2014), (2016).

[6] Throughout this paper "better" means "better, all things considered" if not otherwise indicated.

[7] Note that this holds for *deontic* views too. Plausible deontic views hold that, when all other moral considerations are equal, individual welfare levels are relevant when considering what population to bring about. For a discussion of deontic population ethics, see Arrhenius (2022), (forthcoming).

is not transitive. Therefore, it is possible that each population in the Sequence Argument is imprecisely equally good as the population that comes before it, even though the last population is *worse* than the first population.

However, Parfit's notion of imprecise equality lacked structure. Hájek and Rabinowicz (2022) improved on Parfit's proposal in this regard. In section III we discuss their argument in detail, but in short, their contribution consists in introducing and formalising a notion of *degrees* of incommensurability. An important benefit of their proposal is that they can explain why people *erroneously* (in Hájek and Rabinowicz's view) judge that each population in the Sequence Argument is better than a previous population, when in fact they are incommensurable.

Although Hájek and Rabinowicz's proposal is a step forward, and may help solve many paradoxes, it can only avoid the Repugnant Conclusion at great theoretical cost. First, as we explain in section IV, there is a sequential argument for the Repugnant Conclusion that uses weaker and intuitively more compelling assumptions than the Sequence Argument, and which Hájek and Rabinowicz's proposal only undermines, in a principled way, by allowing for seemingly implausible weight to be put on the disvalue of inequality. Second, if Hájek and Rabinowicz do put such seemingly implausible weight on the disvalue of inequality, then they will have to accept that a population A is not worse than another population B even though everyone in B is better off than anyone in A. So, their proposal then violates the Pareto principle even when the population is held fixed, and thus faces the 'levelling down objection' (Parfit 1995).

In a sense, what we are pointing out is not in any way surprising: one cannot avoid the Repugnant Conclusion without having to accept some counterintuitive implication or make some intuitively implausible assumption. That has been known for decades; hence, the Repugnant Conclusion is often seen as a *paradox* of population ethics. However, what we take to be interesting about the above result is that in order to avoid the Repugnant Conclusion in a principled way, Hájek and Rabinowicz have to violate a *fixed-size* population condition that most would want to accept, namely, the Pareto principle. Giving up the Pareto principle is a pretty hefty price to pay to avoid the Repugnant Conclusion, and Hájek and Rabinowicz have not, as far as we can tell, given us an independent justification for giving up that principle, rather than, say, giving up avoidance of the Repugnant Conclusion.

2. The Sequence Argument for the Repugnant Conclusion and Parfit's response

Consider first the following condition:

> *Quantity*: For any pair of positive welfare levels, 𝔸 and 𝔹, such that 𝔹 is slightly lower than 𝔸, and for any number of lives *n*, there is a greater number of lives *m*, such that a population of *m* lives at level 𝔹 is better than a population of *n* lives at level 𝔸, other things being equal.[8]

Quantity has some intuitive plausibility and should appeal to those who find some truth in the saying "the more good, the better". However, it implies the Repugnant Conclusion together with a reasonable assumption about the structure of welfare:[9]

> *Finite Fine-grainedness*: There exists a finite sequence of slight welfare differences between any two welfare levels.

The idea here is that one can get from one welfare level to another in a finite number of steps of intuitively slight welfare differences. Examples of such welfare differences could be some minor pain or pleasure or a shortening of life by a minute or two.[10] These differences don't have to be of the same size or type. Let's say that a life of type *a* has higher welfare than a life of type *b*, and suppose that you are successively making *a* slightly worse, perhaps by shortening it by a minute or two or by adding some minor pain. Finite Fine-grainedness implies that there is a finite (but possibly great) number of such slight worsening from *a* to another type of life *c* such that a life of this type will have the same welfare or lower welfare than a life of type *b*. It is quite hard to deny the intuitive force of this assumption.[11]

Consider the following sequence of populations for an informal demonstration that these two conditions together imply the Repugnant Conclusion:[12]

---

[8] A welfare level is an equivalence class on the set of all possible lives with respect to the relation "has at least as high welfare as". For an exact statement of this principle, see Arrhenius (2000b), (forthcoming) where this condition is formulated in terms of "at least as good as".

[9] It also implies, and thus presupposes, *the No-Limit Assumption*: For any possible population consisting of lives with a certain welfare, there is a larger possible population consisting of lives with the same welfare. For a discussion, see Arrhenius (2000b), (forthcoming).

[10] For a precise definition of "slight welfare difference" see Arrhenius (forthcoming).

[11] Notice that Finite Fine-grainedness doesn't imply that all sequences of slight welfare differences between two welfare levels are finite, just that there exist at least one such sequence. It is compatible with the welfare ordering being continuous as well as discreet. It just rules out that there are, so to speak, big "jumps" or "holes" in the order of welfare levels. For a discussion of Finite Fine-grainedness and possible theories of welfare that violate this condition, see Arrhenius (2005), (forthcoming); Arrhenius & Rabinowicz (2015). For an interesting effort to challenge Finite Fine-grainedness (in light of the impossibility theorems in population ethics), see Thomas (2018) and Carlson (2022).

[12] For a proof, see Arrhenius (2000b), (forthcoming).

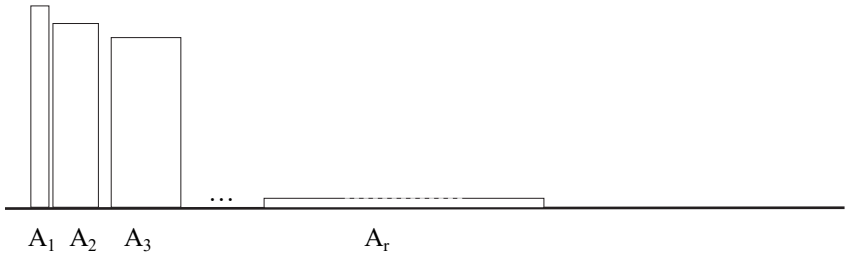


Diagram 2: The Sequence Argument

Assume that A_1 in the diagram above is a population with very high welfare and that A_r is a population with very low positive welfare (again, the width of the blocks represents the number of lives in the population, the height represents their lifetime welfare; dashes indicates that the block in question is much wider than shown). According to Quantity, there is a population A_2 with slightly lower welfare than A_1 and which is better than A_1; a population A_3 with slightly lower welfare than A_2 and which is better than A_2; and so forth. We can assume that the welfare levels in this sequence of populations satisfy Finite Fine-grainedness. Hence, we will finally reach population A_r with very low positive welfare. By transitivity, A_r is better than A_1. Since A_1 is an arbitrary population with very high welfare, this shows that for any population with very high welfare, there is a population with very low positive welfare which is better, that is, the Repugnant Conclusion. Consequently, assuming Finite Fine-grainedness, any theory which avoids the Repugnant Conclusion has to violate Quantity.

As previously mentioned, Parfit (2016) suggests a way of avoiding the sequence derivation of the Repugnant Conclusion by introducing what he calls "imprecision" in value comparisons.[13] He suggests that in a range of important cases, outcomes are only imprecisely comparable. In such cases, transitive relations such as "equally as good as" are not applicable. Instead, we have to make use of imprecise concepts that are non-transitive. This imprecision is not due to any cognitive or epistemic limitations but a fact about the value comparisons of certain types of outcomes.

In the Sequence Argument, Parfit suggested that each population is "imprecisely equally good" to adjacent populations in the sequence. However, since imprecisely equally good is not a transitive relation, he could still maintain that the last population in the sequence is worse than the first population in the sequence. In other words, he had an answer to the Sequence Argument for the Repugnant Conclusion.

[13] Parfit (2014), (2016). Here we are just summarizing his argument, drawing on Arrhenius (2021) where a detailed discussion can be found, to contrast it with Hájek and Rabinowicz theory.

Our aim in this article is not to assess how plausible Parfit's answer was (for an assessment of that, see Arrhenius (2021)). Instead, we shall assess Hájek and Rabinowicz's improvement on Parfit's reply, to which we now turn.

3. Hájek and Rabinowicz's improvement on Parfit's response

Hájek and Rabinowicz's basic observation is that cases that involve incommensurability can differ in *how far from* comparable the relevant options are:

> Sometimes, when attempting to compare two alternatives, we are totally flummoxed, regarding them as not really comparable at all. In other cases, we are more inclined to form a preference one way or another, or to regard them with indifference, but we do so with some hesitancy. And in many of these cases, the hesitancy comes in degrees because incommensurability comes in degrees. (2022: 899)

So, contrary to what Parfit's remarks may have suggested, (in)comparability is not a binary—an either/or—property. Sometimes two options are really incomparable, and sometimes they are really comparable.[14] But sometimes they are somewhere in between, say, close to being comparable. Hájek and Rabinowicz's illustrate their idea with the following example:

> Who was more of a genius: Einstein or Bach? Plausibly, they are incommensurable—one was a great scientist, the other a great composer. How about Einstein or Chopin? Plausibly, they are still incommensurable, but perhaps it is easier to favor Einstein: while Chopin was undoubtedly a genius of piano composition, he arguably did not quite have Bach's range. How about Einstein or Schumann? This comparison is arguably easier again—while brilliant, Schumann was not quite as original as Chopin, let alone Bach. How about Einstein or Salieri, the mediocre composer made famous by Amadeus? That's easy—Einstein was the greater genius, period. We have proceeded by steps to closer and closer approximations to the 'better' relation with regard to genius. (ibid)

Hájek and Rabinowicz's focus is on value comparisons, analysed in terms of *fitting attitudes* (Brentano 1969/1889). On this view, alternative A is better than B if it is *fitting to prefer* A to B, which is taken to mean that one *ought* to prefer A to B. A and

[14] We take it that Hájek and Rabinowicz are here not referring to our abilities to compare, even though their choice of terminology admittedly suggests otherwise, but rather whether the options are in fact comparable. (Thanks to [blinded] for making us see the need to clarify this.)

B are equally good, however, if it is fitting to be indifferent between them, which again means that one *ought* to be indifferent between them. But sometimes, Hájek and Rabinowicz suggest, there may be more than one fitting attitude one could have when comparing A and B. In other words, there could be more than one permissible preference ordering of A vs. B. It might be permissible to rank A over B, and it might also be permissible to rank B over A (or to be indifferent between them). In that case, A and B are *incomparable*, since they contain (or realise) *incommensurable* values.

Given the above understanding of incommensurability, there is a natural way of conceptualising *degrees* of incommensurability:

> We now add that the degree of commensurability can be higher or lower depending on the extent to which different permissible orderings agree or disagree in their ranking of the items. If in nearly all permissible orderings A and B are ranked in the same way, their degree of commensurability is very high—for example, if A is almost always ranked above B, or they are almost always equal-ranked. But if there is more divergence in how A and B are ranked, their degree of commensurability is lower. (Equivalently, their degree of incommensurability is higher.) (2022: 900)

Hájek and Rabinowicz add that if almost all permissible preference rankings of A vs. B have A higher than B, then A is *almost better* than B. In that case, A and B are commensurable to a high degree, but still incommensurable as long as *some* permissible preference raking has B higher than A.

Hájek and Rabinowicz suggest ways of making these degrees precise; most simply, in the finite case, one can simply equate degrees with proportions. The exact details of Hájek and Rabinowicz's proposal are however not all relevant for our purposes. What is relevant is how they apply their general idea to counter the Sequence Argument for the Repugnant Conclusion, while at the same time adding important details to Parfit's similar argumentative structure. As Parfit, Hájek and Rabinowicz suggest that it is false that each population in the Sequence Argument is better than its immediate predecessor. Instead, they are incommensurable. And unlike the better-than relation, the incommensurable-to relation is not transitive. Thus, the Sequence Argument for the Repugnant Conclusion is undermined.

In addition, however, Hájek and Rabinowicz suggest that each population is *almost better* than its immediate predecessor. That would explain why so many people get 'tricked' by the Sequence Argument into endorsing the Repugnant Conclusion, and why very few people say that some (or all) populations in the Sequence Argument are *worse* than their immediate predecessor. So, unlike Parfit, Hájek and Rabinowicz can offer an *error theory* of people's judgement.

> Each [population] is not better than its predecessor, but it is almost better. In fact, it is so close to being better that we mistake the one relation for the other. We do not notice or we ignore the reasonable weighings that do not favor the second population over the first, because they are overwhelmed by those that do. But it is a minor mistake: *almost better* is almost *better*! Our intuitions are wrong, but almost right. This is the error theory that Parfit needed. (2022: 904)

An important question that the above remarks raise is how one should *choose* when one option is *almost better* than another. It does not seem implausible that if, say, A is better than B according to all permissible preference rankings except one, then we ought to choose A over B. But that would mean that Hájek and Rabinowicz cannot avoid a *deontic* version of the Sequence Argument for the Repugnant Conclusion, that is, an argument that is formulated in terms of 'more choiceworthy than' rather than in terms of 'better than'.

Nevertheless, we grant that Hájek and Rabinowicz have suggested an important improvement on Parfit's response to the Sequence Argument. Moreover, the notion of degrees of incommensurability is fruitful outside of population ethics, for instance, promising to solve—or shed light on—paradoxes and puzzles in other areas of philosophy. Unfortunately, however, Hájek and Rabinowicz's proposal can only avoid the Repugnant Conclusion at considerable cost. To appreciate these costs, it is helpful to consider a different (and, in our view, more convincing) sequential argument for the Repugnant Conclusion.

4. The cost of Hájek and Rabinowicz's attempt to avoid the Repugnant Conclusion

Now instead of the sequence in the original Sequence Argument, consider the following:

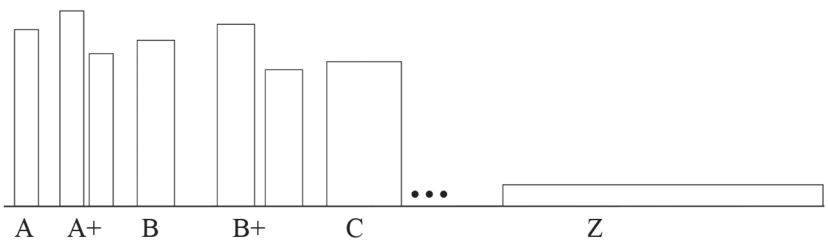


Diagram 3: The Sequential Dominance Addition Argument

All the lives in population A in the diagram above enjoy very high welfare. In A+, we have a collection of lives that is equally large as the collection of lives in A but they enjoy even higher welfare than those in A.[15] In addition, A+ contains a second collection of lives with positive welfare a bit lower than those in A. However, we assume that the welfare of the better-off lives in A+ is sufficiently high to make the average welfare in A+ greater than that in A. It seems to us hard to deny that A+ is better than A, and determinately so. In B, which is of the same size as A+, we have equalized the welfare at a level higher than the +-lives but lower than the A-lives, in a way that increases aggregate (and thus also average) welfare. Unless one has anti-egalitarian intuitions, it seems hard to deny that B is better than A+. And similarly for other consecutive populations in this sequence. But then we are again faced with the Repugnant Conclusion: Z is better than A.

In a moment we will explain the cost of introducing incommensurability to undermine the above "Dominance Addition" argument for the Repugnant Conclusion. But first, let's make the argument more precise, by introducing the two conditions that we implicitly appealed to above when deriving the Repugnant Conclusion. Here's the first one:

> *Dominance Addition*: An addition of lives with positive welfare and an increase in the welfare in all the lives in the rest of the population makes the population better, other things being equal.[16]

One way to motivate Dominance Addition is that you don't make a population worse by adding lives worth living, so if in addition everyone in the new population has higher welfare than anyone in the old population, then you get a better population.

One could make Dominance Addition even more compelling by assuming that the non-added people are the same in the two compared populations. Then one could also appeal to so-called *person-affecting view* for judging A+ better than A since then the A-people will benefit in the move from A to A+. We shall not avail ourselves of this possibility here, however, since the person-affecting view has been shown to be deeply problematic for many reasons. We shall continue to assume that the compared populations are pairwise disjoint. Those who still think the person-affecting view can be salvaged may however make that assumption which some will find strengthens the intuitive appeal of Dominance Addition.

---

[15] Notice, as we stated in fn. 2, that populations represented by different letters, or the same letter but different indexes, are pairwise disjoint.

[16] For an exact statement of this condition, see Arrhenius (2000b), (forthcoming) where it is formulated in a logically weaker manner in terms of "not worse than". We are using the stronger formulation here to simplify the exposition.

Dominance addition is an intuitively more compelling version of the more well-known *Mere Addition Principle*: An addition of people with positive welfare does not make a population worse, other things being equal.[17] Yet, although this principle might seem a compelling principle at first glance, it is controversial. Several authors have rejected it.[18] One might, for example, object to it on egalitarian grounds since a mere addition can introduce great inequality in an otherwise perfectly equal population.[19] Likewise for Dominance Addition albeit then the disvalue of the introduced inequality also has to be weighed against the positive value of the increased welfare of the lives in the original population, not only against the possible positive value of more lives with positive welfare. We shall get back to such objections to Dominance Addition in a moment. But first, we introduce the second condition we appealed to informally above when deriving the Repugnant Conclusion:

> *Inequality Aversion*: For any triplet of welfare levels, $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{C}$, $\mathbb{A}$ higher than $\mathbb{B}$ and $\mathbb{B}$ higher than $\mathbb{C}$, and for any population A with welfare $\mathbb{A}$, there is some larger population C with welfare $\mathbb{C}$ such that a perfectly equal population B of the same size as A∪C and with welfare $\mathbb{B}$ is better than A∪C, other things being equal.[20]

Another way of stating Inequality Aversion is that for any welfare level of the best off and worst off, and for any number of best off lives, there is some (possibly much) greater number of worst off lives such that it would be better to have an equal distribution of welfare on any level higher than the worst off, other things being equal.

The above is a very weak egalitarian condition since it can be satisfied by a theory which demands that the total welfare must be greater for a population with perfect equality to be better than an unequal population of the same size. Moreover, it is also compatible with principles that give much greater weight to the welfare of the best off as compared to the welfare of the worst off. For example, a theory which requires that to compensate for one life falling from twenty to ten units of welfare, a hundred

---

[17] Cf. Parfit (2014), p. 420ff, Hudson (1987), Ng (1989), and Sider (1991). Cf. fn. below. Notice that the original formulation of this condition in Arrhenius (2000b), (forthcoming) is also logically weaker than the Mere Addition Principle.

[18] Ng (1989), p. 244; Blackorby, Bossert, & Donaldson (1995), p. 1305, and Blackorby, Bossert, & Donaldson (1997), pp. 210–211; Fehige (1998). Ng ascribes to Parfit the view that a population axiology should satisfy the Mere Addition Principle (Ng (1989), p. 238) and one might get that impression from Parfit (2014), p. 420ff. In personal communication, however, Parfit has expressed doubts about the Mere Addition Principle in cases where the added people are much worse off than the rest of the population. See also Feldman (1997) ch. 10, Kavka (1982), and Carlson (1998), pp. 288–289.

[19] See Arrhenius (2009), (2013), (forthcoming).

[20] For an exact statement of this principle, see (2000b), (forthcoming) where this condition is formulated in terms of "at least as good as". We've here formulated it in terms of "better than" to simplify the exposition.

lives have to be moved from zero to ten units, is compatible with Inequality Aversion. In that sense, its name is a bit misleading since it is compatible with quite non-egalitarian theories. Roughly, Inequality Aversion only rules out theories that imply that we should always or sometimes give some kind of "lexical priority" to the best off.[21] A simple example of such a theory is "Maximax": Maximise the welfare of the best off.

Let's return to diagram 3. Dominance Addition implies that A+ is better than A. We can assume that A+ and B fulfil the antecedent of Inequality Aversion.[22] So, Inequality Aversion implies that B is better than A+. Likewise for populations B, B+, and C, and so forth until we finally reach population Z with very low positive welfare. By transitivity, Z is better than A, that is, the Repugnant Conclusion.

Now, it does not seem to us that Hájek and Rabinowicz's proposal gives us resources to deny Inequality Aversion. For instance, we can assume that everyone's lives in both the A+ world and the B world contain the 'best things in life' (cf. Parfit (1986), (2016)). Moreover, we can assume each life in B contains the same quality and amount of the best things in life as each life in A+, it is just that the bad things (pain and suffering, etc.) are more equally distributed in B than in A+. Now, some might object that although this may be plausible for A+ and B, it is less plausible that once we get further down the sequence (towards lives barely worth living), it will still be true that all lives in the worlds we are comparing contain the same amount and quality of the best things in life. However, since we are concerned with *lifetime* welfare, when evaluating whether a life is, say, barely worth living, we don't see any principled reason for why all lives in the Dominance Addition Sequence couldn't contain the same quality and amount of the best things in life. After all, we can, for instance, simply imagine extending the lives, but adding to them more and more suffering (or simply longer and longer very boring periods). So, concern for the 'best things in life' does not, we think, undermine Inequality Aversion (for further discussion of this issue, see Arrhenius 2021).

So, let's suppose instead that Hájek and Rabinowicz want to resist the Sequential Dominance Addition Argument by rejecting Dominance Addition. They do in fact have the formal resources to do so. For they could claim that there is a permissible preference ordering that ranks A above A+, for instance, a preference ordering that puts very high weight on the disvalue of inequality. (In a moment we shall consider another reason for why there could be a permissible preference that ranks A above

[21] There are some more subtle theories that violate Inequality Aversion, such as theories that invoke some form of superiority in value. See Arrhenius (2005); Arrhenius & Rabinowicz (2005), (2015) for a discussion. As we shall discuss below, Inequality Aversion can be derived from an even more intuitively compelling condition, Non-Elitism.

[22] If welfare is measurable on at least an interval scale, we could also assume that the total and average welfare in B is higher than in A+.

A+.) However, the preference in question really would have to put *a lot* of weight on the disvalue of inequality. After all, we can make the population that gets better lives when we move from A to A+ arbitrarily large, and we can similarly make the additional people in A+ (whose lives are worth living) arbitrarily numerous.[23] So, to avoid saying that A+ is determinately better than A by appealing to the permissibility of valuing equality, Hájek and Rabinowicz have to say that it is permissible to give what seems to us to be implausibly high importance to equality. And while their framework makes room for such judgements, nothing in their paper gives us *good reasons* for such judgements. Let's however set that issue aside, and consider another issue that now arises.

Consider diagram 4. We assume that the number of people in A' is *n*, which is the same as the number of the worse-off people in A'+. The *n* worse-off people in A'+ are better off than the people in A'. In addition, A'+ contains some even better off people. Population B' however contains exactly the same number of people as population A'+, but in B' everyone is worse off than the worse-off people in A+ but still better off than the people in A'.

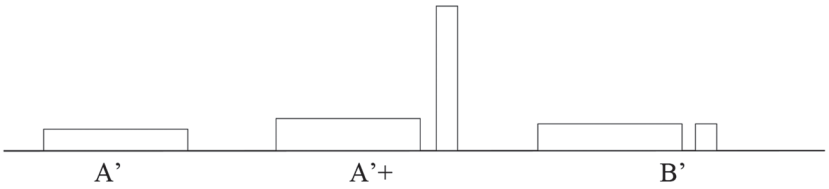


Diagram 4: Levelling down

Now compare population A' with population A'+. Here it would seem that Hájek and Rabinowicz would have to say that the latter is only *almost* better than the former; that is, there is some permissible preference according to which A' ranks higher than A'+, namely, a preference that places a very high weight on the disvalue of inequality. At the very least, there will have to be *some* similar pair of populations for which they will have to say that the population containing both more people and higher welfare for everyone is only *almost* better, if they are to resist the Sequential Dominance Addition Argument for the Repugnant Conclusion by claiming that A+ is not determinately better than A due to the added inequality in the former.

[23] We are assuming that Hájek and Rabinowicz do not deny that the number of people enjoying very high levels of welfare is of *some* moral importance. After all, if they denied that, say, total welfare is of *any* moral importance, then that would suffice to block the Sequence Argument (without appealing to incommensurability).

What about A' vs. B'? It is hard to see how there could be a permissible preference that does not rank B' over A'. The difference between the two is that, first, everyone in B' is better off than anyone in A', and, second, B' contains more people with lives worth living. But there is no added inequality in B' compared to A'; nor is there anything else in B' but not in A' that could, in our view, plausibly be of negative value. So, if *either* having more people with lives worth living makes a world at all better, no matter how slight, *or* if everyone being better off makes a world at all better, then we must say that B' is better than A'. For the purposes of our argument, it however suffices that B' is at least as good as A' (as should be apparent below).

In response to the last paragraph, some might point out that there is a respectable view according to which B' does contain something of negative value that A' does not. For according to *Critical Level Utilitarianism* (CLU), adding lives with positive welfare under a positive critical level has negative value. So, if the people in both A' and B' are below the critical level, then the fact that there are *more* people in B' might make the former better, according to CLU. Two things could be said in response.

First, if A' is worse than B', due to the aforementioned reason, then we can instead focus on different populations A'' and B'' that differ from A' and B' in that the number of people that are common to both populations is much greater in A'' and B'' than in A' and B'. For some such pair of populations, A'' and B'', we should find that B'' is better than A'' according to CLU, even though B'' contains more people below the critical level than A'', since B'' brings so many people *closer to* the critical level.

Second, and maybe more importantly given the present argument, Hájek and Rabinowicz can hardly appeal to CLU in response to our argument. The reason is that if a critical level is allowed, then we already have a response to the Sequence Argument, since once we get below the critical level in the sequence, the populations get worse and worse, according to CLU, the further along the sequence we go. Hence, Hájek and Rabinowicz's proposed solution would be superfluous. This remark of course holds more generally: we assume that any view or principle that Hájek and Rabinowicz might want to invoke in response to our argument should not make their response to the Sequence Argument superfluous. (Finally, it may be worth mentioning that CLU violates Non-Sadism and other plausible adequacy conditions (Arrhenius (2000a), (2000b), (forthcoming)).

So, we can safely assume that Hájek and Rabinowicz won't respond to our argument by assuming CLU. Is there some other way to deny that the claim that B' is at least as good as A' (in Diagram 4)? Perhaps the most principled way to deny that claim, we think, is to say that populations of different sizes are *always* incommen-

surable. In fact, Parfit briefly considered such a view.[24] That however seems to us very implausible (and, in fact, Parfit himself abandoned the view). For instance, it would imply that a population in Stone Age conditions, where nobody has an excellent life and most people lead very miserable lives, is no worse than a greater population in which a huge number of people live in great luxury thanks to technological and moral advancement.[25]

We can thus assume that B' is at least as good as A'. However, recall that to avoid the Repugnant Conclusion, Hájek and Rabinowicz have to say that A'+ is merely *almost* better than A'. Therefore, since better-than is a transitive relation, they have to deny that A'+ is better than B'**.** But that seems counterintuitive (even if they can say that A'+ is *almost* better than B). These populations contain the same number of people, but everyone in A'+ is better off than anyone in B'. In fact, some people in A'+ are *much* better off than anyone in B'. (Those with a strong aversion to inequality could however diminish the gap between the better of and the worse off in A'+. It would of course still be the case than everyone in A'+ is better off than anyone in B'.) So, Hájek and Rabinowicz have to reject a weak version of the widely endorsed Pareto principle for fixed-sized populations, according to which a population A* is better than an equi-sized population B* if everyone in A* is better off anyone in B*. For the same reason, they face the levelling down objection.

Is there some way for Hájek and Rabinowicz to resist the above implication while also resisting the Sequential Dominance Addition Argument for the Repugnant Conclusion? We can think of one response on their behalf. They could argue that the reason A+ is not determinately better than A is that there is a preference that ranks A over A+, but not in virtue of the inequality in the latter, but rather because in the latter it is not true that *everyone has a fantastic life*. At least, that would plausibly be true for some pair of worlds with the relevant relationship, that is, where one is a "dominance addition" of the other. But if they claim that it is *not* permissible to base one's preference for A over A+ on concern for equality, then they don't have to say that it is permissible to prefer A' over A'+; so, they don't have to violate the Pareto principle.

But is the above response plausible? We think not. To resist the Sequential Dominance Addition Argument for the Repugnant Conclusion, Hájek and Rabinowicz have to be very liberal about what can be permissibly preferred and what reasons one can permissibly have for one's preferences. In particular, they have to say that it is permissible to prefer A over A+ *because* only in the former world does everyone have a fantastic life. (Or at least, they have to say that of some worlds where the latter

[24] See in particular Parfit's Rolf Schock Prize Lecture and his unpublished 2014 manuscript based on the lecture. See also Arrhenius (2016) for a lengthier discussion of this view.

[25] Thanks to [blinded] for suggesting to us an example like this.

is a dominance addition of the former.) But in another sense, they cannot be liberal about what can be permissibly preferred: they have to say that it is impermissible to prefer A over A+ *because* the latter contains inequality.

The above response that we are considering on Hájek and Rabinowicz's behalf therefore strikes us as being rather odd. Inequality is a widely recognised value and many people think it is fitting to accept considerable cost to bring about inequality. But the same doesn't seem true about everyone having fantastic lives. There is, for instance, no traditional distributive view that places a particular significance on *everyone* having fantastic lives. Egalitarians think that it is good that everyone is equally well-off; but if that justifies preferring A over A+, then that is because of the importance of *equality*, not because of the importance of everyone having fantastic lives. Utilitarians by contrast place greater weight on everyone having fantastic lives than on equality; but utilitarian principles do not justify preferring A over A+. More generally, it seems to us that it would be hard to find a principled and ethically sound justification for preferring A over A+ that is not grounded in the value of equality. But then it may not be possible to satisfy the Pareto principle and avoid the levelling down objection.

5. Concluding remarks

Before concluding, we would like to acknowledge again that, first, Hájek and Rabinowicz's proposal is interesting in its own right and may shed light on various paradoxes in philosophy; and, second, that their response to the Sequence Argument is an improvement on Parfit's. Nevertheless, their proposal can only help us avoid the Repugnant Conclusion at great cost. For as we have now demonstrated, it seems that the only principled way in which their proposal can avoid the Repugnant Conclusion is by allowing the desire to avoid inequality to play a seemingly implausibly strong role; so strong that we would sometimes have to say that one population is no better than another population even though everyone in the one population is better off than anyone in the other population. In other words, they violate the Pareto principle and thus face the levelling down objection. This is a pretty hefty price to pay in order to avoid the Repugnant Conclusion.

References

Arrhenius, G. (2000a). An Impossibility Theorem for Welfarist Axiologies. *Economics and Philosophy*, *16*(02), 247–266.

Arrhenius, G. (2000b). *Future Generations: A Challenge for Moral Theory*. Retrieved from http://www.diva-portal.org/smash/record.jsf?pid=diva2:170236

Arrhenius, G. (2005). Superiority in Value. *Philosophical Studies*, *123*(1/2), 97–114. https://doi.org/10.1007/s11098-004-5223-0

Arrhenius, G. (2009). Egalitarianism and Population Change. In A. Gosseries & L. Meyer (Eds.), *Intergenerational Justice* (1st Edition, pp. 325–349). Oxford: Oxford University Press.

Arrhenius, G. (2011). The Impossibility of a Satisfactory Population Ethics. In H. Colonius & E. N. Dzhafarov (Eds.), *Descriptive and Normative Approaches to Human Behavior, Advanced Series on Mathematical Psychology* (pp. 1–26). World Scientific Publishing Company.

Arrhenius, G. (2013). Egalitarian Concerns and Population Change. In O. Frithjof Norheim, N. Eyal, S. A. Hurst, & D. Wikler (Eds.), *Inequalities in Health: Concepts, Measures, and Ethics* (pp. 74–91). Oxford: Oxford University Press.

Arrhenius, G. (2016). Population Ethics and Different-Number-Based Imprecision. *Theoria*, *82*(2), 166–181. https://doi.org/10.1111/theo.12094

Arrhenius, G. (2021). Population Ethics and Conflict-of-Value Imprecision. In J. McMahan, T. Campbell, & J. Goodrich (Eds.), *Ethics and Existence: The Legacy of Derek Parfit*. Oxford: Oxford University Press.

Arrhenius, G. (2022). Population Paradoxes without Transitivity. In G. Arrhenius, K. Bykvist, T. Campbell, & E. Finneron-Burns (Eds.), *The Oxford Handbook of Population Ethics* (pp. 180–203). https://doi.org/10.1093/oxfordhb/9780190907686.013.11

Arrhenius, G. (forthcoming). *Population Ethics: The Challenge of Future Generations*. Oxford University Press.

Arrhenius, G., & Rabinowicz, W. (2005). Millian Superiorities. *Utilitas*, *17*(2), 127–146. https://doi.org/10.1017/S0953820805001494

Arrhenius, G., & Rabinowicz, W. (2015). Value Superiority. In I. Hirose & J. Olson (Eds.), *The Oxford Handbook of Value Theory* (pp. 225–248). New York NY: Oxford University Press.

Blackorby, C., Bossert, W., & Donaldson, D. (1995). Intertemporal Population Ethics: Critical-Level Utilitarian Principles. *Econometrica*, *63*(6), 1303–1320. https://doi.org/10.2307/2171771

Blackorby, C., Bossert, W., & Donaldson, D. (1997). Critical-Level Utilitarianism and the Population-Ethics Dilemma. *Economics and Philosophy*, *13*(02), 197–230. https://doi.org/10.1017/S026626710000448X

Broome, J. (1999). *Ethics out of Economics*. Cambridge: Cambridge University Press.

Broome, J. (2004). *Weighing Lives*. Oxford: Oxford University Press.

Bykvist, K. (2007). The Good, The Bad, and the Ethically Neutral. *Economics and Philosophy*, *23*(01), 97–105. https://doi.org/10.1017/S0266267107001253

Carlson, E. (1998). Mere Addition and Two Trilemmas of Population Ethics. *Economics and Philosophy*, *14*(02), 283–306.

Carlson, E. (2022). On Some Impossibility Theorems in Population Ethics. In G. Arrhenius, K. Bykvist, T. Campbell, & E. Finneron-Burns (Eds.), *The Oxford Handbook of Population Ethics* (1st ed., pp. 204–225). https://doi.org/10.1093/oxfordhb/9780190907686.013.14

Fehige, C. (1998). A Pareto Principle for Possible People. In C. Fehige & U. Wessels (Eds.), *Preferences* (pp. 508–543). Berlin ; New York: W. de Gruyter.

Feldman, F. (1997). *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*. Cambridge: Cambridge University Press.

Hudson, J. L. (1987). The Diminishing Marginal Value of Happy People. *Philosophical Studies*, *51*(1), 123–137. https://doi.org/10.1007/BF00353967

Kavka, G. S. (1982). The Paradox of Future Individuals. *Philosophy & Public Affairs*, *11*(2), 93–112.

Ng, Y.-K. (1989). What Should We Do About Future Generations? *Economics and Philosophy*, *5*(02), 235–253.

Parfit, D. (1984). *Reasons and Persons* (1991st ed.). Oxford: Clarendon.

Parfit, D. (1986). Overpopulation and the Quality of Life. In P. Singer (Ed.), *Applied Ethics* (1 edition, pp. 145–164). Oxford: New York: Oxford University Press.

Parfit, D. (2014). *How We Can Avoid the Repugnant Conclusion*. Mimeo, University of Oxford, Faculty of Philosophy.

Parfit, D. (2016). Can We Avoid the Repugnant Conclusion? *Theoria*, *82*(2), 110–127. https://doi.org/10.1111/theo.12097

Sider, T. R. (1991). Might Theory X Be a Theory of Diminishing Marginal Value? *Analysis*, *51*(4), 265–271.

Thomas, T. (2018). Some Possibilities in Population Axiology. *Mind*, *127*(507), 807–832. https://doi.org/10.1093/mind/fzx047

Katharina Berndt Rasmussen[1]

Discrimination and Future Generations[2]

In this paper, I analyse whether the present generation's choices to, e.g., deplete resources, shift environmental burdens towards the future, and discount the lives and interests of future generations, can be instances of *discrimination against future generations*. This has been tentatively suggested in both legal theory and philosophy; I review such suggestions briefly in section 1. However, a more rigorous analysis – outlining the concept, relevant grounds, and wrong-making features of discrimination, and applying these to future generations – is still lacking. To address this lacuna, I propose a theory of discrimination and analyse why it might seem to apply – yet ultimately fails to apply – to the differential treatment of future generations. More specifically, I propose a definition of discrimination (section 2.1) and an account of the moral wrongness of discrimination (section 2.2). I moreover explore the connection between discrimination and theories of social (in)justice (section 2.3). I then apply this theory to the problem of differential treatment of future generations. While discrimination may occur between collectives, such as generations (section 3.1),

[1] Stockholm University & Institute for Futures Studies, katharina.berndt@philosophy.su.se

[2] I would like to thank the audiences of the Final conference of the Climate Ethics and Future Generations program (IFFS Stockholm, September 2023), the Gothenburg workshop with Jennifer Lackey (Philosophy Department, Gothenburg University, October 2023), the CEPDISC workshop on Structural Discrimination (Aarhus University, October 2023), the Higher Seminar in Practical Philosophy (Philosophy Department, Stockholm University, November 2023), and the 8th Stockholm Graduate Conference (Philosophy Department, Stockholm University, December 2023), for helpful comments on drafts of (different parts of) this paper. Generous funding by Riksbankens Jubileumsfond and the Swedish Research Council are gratefully acknowledged.

my analysis shows that the specific temporal status of future generations is not comparable to other grounds of discrimination, such as gender or race (section 3.2). Moreover, due the non-identity problem and the problem of lack of a "community of social meaning" between generations, future generations cannot be claimed to be subjected to worse treatment by the present generation (section 3.3). Hence, their differential treatment due to the present generation's choices does not amount to discrimination. Section 4 concludes and outlines some upshots of my analysis.

1. Introduction

The recently adopted *Maastricht Principles on The Human Rights of Future Generations* state, under the heading of §I.6. Equality and Non-Discrimination:

> "Future generations must be free from intergenerational discrimination. This discrimination includes but is not limited to:
>
> i. The waste, destruction, or unsustainable use of resources essential to human life;
> ii. Shifting the burden of responding to present crises to future generations; and
> iii. According less value to future lives and rights than the lives and rights of present generations, including discounting the impacts and burdens of present conduct on the lives and rights of future generations."[3]

The *Maastricht Principles*, which are based on the United Nations report *Our Common Agenda*, develop human rights standards to increase the protection of the human rights of future generations. They aim to guide political and legal institutions, as well as social movements, on regional, national, and international levels.[4] The principles define future generations as "those generations that do not yet exist but will exist and who will inherit the Earth [which includes] persons, groups and peoples" (§ I.1). Though the principles are framed mainly in terms of the human rights of future generations, they do contain a few paragraphs specifically on discrimination.[5] However, they do not spell out what exactly is meant by 'discrimination', neither legally nor colloquially speaking.

[3] Adopted at the *Maastricht Centre for Human Rights*, on 3 February 2023; https://www.rightsoffuturegenerations.org/the-principles.

[4] (Franco and Liebenberg 2023).

[5] The *Maastricht Principles* also contain paragraphs applying to states' obligations of non-discrimination: "Violations of [state] obligations to respect the human rights of future generations include [...]: Engaging in conduct that results in discriminatory access to natural resources and benefits enjoyed by future generations as compared to present generations" (§ II.17).

The idea of discrimination against future generations has long roots in legal theory. Published 1978, "Discrimination against Future Generations" is one of the first articles that systematically examine the US Constitution's provision to "Posterity". Legal scholar Jim Gardner argues that this provision implies a policy of intergenerational fairness which may "in certain circumstances limit the power of state and federal governments to impose disadvantages on future generations".[6] The article makes frequent reference to the US Constitution's fourteenth amendment, which includes the provision of *equal protection of the laws* for all US citizens. Recent work in US legal theory keeps up this focus on the fourteenth amendment, mandating non-discrimination, in light of the ever more urgent climate crisis.[7]

However, among recent climate lawsuits by young activists against states (such as the US, Sweden, Germany, the Netherlands), none have invoked discrimination law as pertaining to future generations, in the sense specified by, e.g., the above Maastricht Principles: as those generations who do *not yet exist*. When discrimination law is invoked in these lawsuits, it concerns discrimination against young – i.e., *existing* – individual complainants or generations. Consider the description of two such cases:

> "The problem of birth cohort discrimination is raised in the currently pending Duarte Agostinho case, initiated by Portuguese children and minors before the ECtHR. The applicants launched their complaint against Portugal and 32 other States for violating Article 2 (right to life) and 8 (right to family life) in conjunction with Article 14 (non-discrimination) of the ECHR. They argue, that due to the respondents' failure to adopt stringent mitigation measures, the *complainants* will experience extreme weather events, including heatwaves, which affect *their* living conditions and health.
> [...]
> Anti-age discrimination claims are also filed with domestic courts. In Canada, the Superior Court of Justice Ontario in Mathur deemed the "adverse effects of climate change on *younger generations*" to be "self-evident" and allowed the claim to proceed to trial."[8]

When it comes to moral and political philosophy, there is a similar fissure. Also in these fields, climate inaction has been discussed under descriptions such as "discrimination between generations",[9] "discrimination by date of birth",[10] and "discrimi-

[6] (Gardner 1978, 33).

[7] See e.g. (Campbell 2019), (Nguyen 2017).

[8] (Sulyok 2023), my italics.

[9] (Attfield 2010).

[10] (Stern 2014).

nation against future generations".[11] However, the lion share of the philosophical debate around intergenerational inequities concerning climate change does not refer to the concept of discrimination.[12] And in the texts that do, the specifics concerning discrimination – as, at a conceptual minimum, *unequal disadvantageous treatment* – of future generations are not spelled out in any detail. That is, there is a lack of a more rigorous analysis, of outlining the concept, relevant grounds, and wrong-making features of discrimination – and applying these to future generations. We are therefore not in a position to determine whether and when discrimination against future generations is instantiated, nor when and why it is wrong.[13] My aim in this paper is to address this lacuna. I propose a theory of discrimination and analyse why it might seem to apply – yet ultimately fails to apply – to the differential treatment of future generations.

I proceed as follows: in section 2, I outline a theory of discrimination with the following steps: I propose a definition of discrimination and show how it gives rise to four distinct forms of discrimination (in 2.1). Then, I propose an account of the moral wrongness of discrimination, employing a broad (unorthodox) concept of harm (in 2.2). I moreover argue that this account still does not exhaustively capture what makes the phenomenon of discrimination problematic. I therefore, finally, explore the connection between discrimination and theories of social (in)justice (in 2.3). In section 3, I apply the theory developed in section 2 to the problem of differential treatment of future generations. I closely analyse four cases, which seem to instantiate the four distinct forms of discrimination, respectively. Yet, while discrimination arguably may occur between collectives, such as generations (in 3.1), my analysis shows that the specific ground of discrimination – temporal (future) status – is not comparable to other grounds of discrimination, such as gender or race (in 3.2). Moreover, due the non-identity problem and the problem of lack of a "community of social meaning" between generations, future generations cannot, in the end, be claimed to be subjected to worse treatment by the present generation (in 3.3). Hence, their differential treatment does not amount to discrimination. Section 4 concludes and outlines some upshots of my analysis.

[11] (Gardiner 2017).

[12] E.g., the search terms "discrimination" AND "future generations" return 13 papers on PhilPapers, while "justice OR injustice" AND "future generations" return 291 papers (as of 2024-01-11).

[13] Gosseries has examined the conditions for successful climate lawsuits in terms of age discrimination within different legal frameworks (Gosseries 2015). While such a legal approach may point towards a pragmatically promising venue towards climate justice, it does not seriously consider the philosophical foundations of such a strategy. I am here interested in precisely this foundation. For my pragmatic response to Gosseries' approach, see section 4 below.

2. A theory of discrimination

So, what is discrimination? In its widest sense, to discriminate is to distinguish or differentiate between things. Obviously, I'm here interested in a much narrower sense. I want to capture the specific phenomenon of discrimination that most of us are concerned with in real life. I propose, tentatively, that this concerns the differential *treatment* of others (paradigmatically: persons), which is in some sense *detrimental* to them, and typically connected to some (perceived) *group* membership. Moreover, instances of the phenomenon typically appear to us as *problematic* (this captures the negative valence contained in utterances such as "But that's discrimination!", or the normative status ascribed to it in legal documents).

2.1. Definition

The following generic definition of (group) discrimination aims to accommodate the above concerns.[14]

> *Definition*: An agent, X, (group) discriminates against someone, Y, in context C by ϕ-ing if and only if:
>
> (i) there is a property, P, such that Y has P (or X believes that Y has P),
>
> (ii) by ϕ-ing, X treats Y worse than X would have treated Y, had Y not had P (or had X not believed Y to have P),
>
> (iii) it is because Y has P (or because X believes that Y has P) that X treats Y worse by ϕ-ing, and
>
> (iv) P is the property of being a member of a socially salient group in C.

This generic definition is meant to capture both direct and indirect discrimination, as commonly understood. In fact, this definition helps expand our conceptual framework to accommodate further forms of discrimination, by bringing out that the orthodox distinction 'direct/indirect' is conflated and in need of clarification. To see this, consider that conditions (ii) and (iii) can each be interpreted in (at least) two different ways. Combining the resulting two distinctions provides a conceptual map of four (rather than two) distinct forms of discrimination.

14 See (Berndt Rasmussen 2019), (Berndt Rasmussen 2020), (Berndt Rasmussen 2023). The definition is close to a number of other definitions in the literature. See e.g. (Lippert-Rasmussen 2014) and many of the entries in (Lippert-Rasmussen 2017).

To start, consider (ii). This condition can be interpreted as:

> (ii') X *would have π-ed,* rather than ϕ-ed, had Y not (been believed to) have P, and ϕ-ing toward someone constitutes worse treatment of them than π-ing,

or as:

> (ii") had Y not (been believed to) have P, X *would still have ϕ-ed,* but ϕ-ing toward someone with P constitutes worse treatment than ϕ-ing toward someone without P.

The first interpretation gives us discrimination as *differential treatment*: comparing two different acts, ϕ-ing vs π-ing. The second gives us discrimination as *disparate impact*: comparing how one "facially neutral" act, ϕ, impacts differently on someone with P vs someone without P.

Even condition (iii) can be interpreted in two alternative ways, as:

> (iii') it is because *X has P-related intentions* (e.g., X dislikes people with P and believes that Y has P) that X treats Y worse,

or as:

> (iii") it is not because X has P-related intentions that X treats Y worse, but rather because of some *other P-related cause*.

The first interpretation gives us *intentional* discrimination: property P plays a motivational role for X's action, by figuring in the content of X's motivating beliefs or desires. The second gives us *non-intentional* discrimination: property P has an explanatory role with regard to X's action, but not by figuring in the content of X's motivating beliefs or desires.

Combining these two pairs of distinctions results in four possible forms of discrimination. *Table 1* systematises these and exemplifies each form with a paradigmatic case of *race discrimination in enrolment/employment decisions*, where a gatekeeper (X) refuses to accept (ϕ) an applicant (Y), making the applicant worse off by denying a sought opportunity, just due to the applicant's being black (property P).

	Differential treatment	Disparate impact
Intentional	(1) A university in the early 1950's US South accepts a white applicant but turns down an *equally* qualified black applicant, stating: "This is a whites-only university. Blacks are referred to apply to some 'separate-but-equal' university for African Americans."[15]	(2) An employer turns down a *qualified* black applicant, stating: "We don't hire people who lack high school education", while intentionally using this criterion because of its ability to track politically induced, race-correlated educational deficits.[16]
Non-intentional	(3) A university accepts a white candidate for their PhD-programme but turns down an *equally* qualified black candidate, ranking the latter as less qualified, where the ranking is due to the evaluators' implicit biases.[17]	(4) An employer turns down a *qualified* black applicant, stating: "We don't hire people who lack high school education", without any awareness of the criterion's ability to track politically induced, race-correlated educational deficits.[18]

Table 1: Four forms of discrimination with paradigmatic examples.[19]

I will, in the rest of this paper, rely on the above generic definition of group discrimination. In section 3, I will return to the four forms of discrimination specified in *Table 1*, in order to analyse the disadvantageous treatment of future generations by the present generation's climate inaction. In the remainder of this section, I will consider the problematic features of discrimination, analysing them in terms of wrongness and injustice.

[15] This example resembles Sweatt v. Painter; see (Lavergne 2010). Note that there may but need not be disparate impact under disparate treatment: if (contrary to historical fact) the educational facilities had been separate and relevantly equal, blacks might not have been worse off than whites in the labour market, but such non-disadvantageous yet differential treatment would still constitute discrimination and may still be marked as morally wrong as such.

[16] This example resembles Griggs v. Duke Power Company; see (Khaitan 2015, 31), but with the addition that the criterion "is covertly used to target members of a protected class" (Mendoza 2017, 258). Cf. even Altman's "Jim Crow era" example (Altman 2016, para. 2.1).

[17] This example may be posited as a specific instantiation of the unequal rankings of identical CVs under different (racially or gender coded) names, which have been extensively studied (Zschirnt and Ruedin 2016). Cf. (Alesina et al. 2018) for the correlation of teachers' implicit anti-immigrant bias and their grading of immigrant vs. native middle school students.

[18] This example resembles Griggs v. Duke Power Company under "absence of a discriminatory intent" (Khaitan 2015, 31). There is, of course, the separate but related problem of discrimination at the educational level.

[19] *Table 1* appears originally in (Berndt Rasmussen 2020, 738).

2.2. Moral wrongness

The definition of discrimination, as it stands, is non-moralised. But it brings out a normative feature that should be captured by any plausible account of the moral wrongness of discrimination: that Y is subjected to *worse treatment* by X, in some sense, as stated by condition (ii). This allows different accounts of moral wrongness to be plugged in here, spelling out worse treatment in terms of, e.g., disrespect,[20] demeaning,[21] freedom violation,[22] or harm.[23] These accounts offer different explanations of the prima facie wrongfulness of discrimination. What they have in common is that they focus on how the discriminatee is wronged by the discriminator's action – they simply differ in spelling out the details of this wronging.[24]

I have, in a previous paper, argued for a counterfactual-harm based account – albeit one that appeals to an unorthodox, broad concept of harm.[25] This concept has a welfarist component (capturing the orthodox sense of harm[26]): by ϕ-ing, X makes Y worse off, i.e., lowers their well-being compared to the counterfactual where Y had not had (or had not been believed to have) P. The broad harm concept also has a non-welfarist component, of being treated as inferior in some sense. This latter component captures features in the vicinity of the above-mentioned disrespect and demeaning accounts. We may here fill it in by using Deborah Hellman's influential account: discrimination is wrong when it is demeaning, in the sense that it, first, "expresses that a person or group is of lower status [...] and, second, the actor or institution expressing this meaning must have sufficient social power for this expression to have [the capacity to have] force".[27]

This account could also be spelled out as a hybrid harm-*and*-inferior-treatment (e.g., demeaning) approach. The issue is mainly terminological, although I believe that there is something to be said for the idea that being treated as inferior can reasonably be seen as a kind of (relational) harm to the individual. In any case, the welfarist and non-welfarist components are meant to capture what is intuitively problematic with paradigmatic cases of discrimination: that people – individually or in groups – suffer from such differential treatment; that it is a wrong that is directed

[20] (Eidelson 2015); cf. (Beeghly 2017).

[21] (Hellman 2008), (Hellman 2017).

[22] (Moreau 2010), (Moreau 2013), (Moreau 2017).

[23] (Arneson 2006), (Lippert-Rasmussen 2014).

[24] Note that these accounts are not mutually exclusive; several of these might be combined to provide hybrid or pluralist accounts of the wrongness of discrimination.

[25] (Berndt Rasmussen 2019)

[26] Cf. (Parfit 1986, 487).

[27] (Hellman 2017, 102); cf. (Hellman 2008).

against them.[28] I will not defend this account here, but merely stipulate it as a basis for further discussions of the structural dimensions of discrimination:

> *Moral wrongness*: An instance of (group) discrimination, ϕ-ing, by X against Y, on grounds of P, is (prima facie) wrong because X by ϕ-ing treats Y worse, in the following sense:
>
> (a) making Y worse off, or
>
> (b) treating Y as inferior (in the sense of demeaning),
>
> than X would have, had Y not had P (or had Y not been believed to have P).[29]

Note that the definition of discrimination, while formulated in evaluative terms of 'treating worse', does not make any moral claim. Such a claim is introduced by the separate and non-definitional account of the wrong-making feature of discrimination. The definition is thus non-moralised. The theory in its entirety of course is not – and should not be. Recall that I set out to capture the specific phenomenon of discrimination that most of us are concerned with in real life: as a phenomenon that typically appears to us as *problematic*. Now, my definition carves out a very specific social phenomenon (group discrimination), which most of us take issue with. The wrongness account then merely spells out why we are morally justified in taking issue with it, i.e., why it is an apt candidate for at least *prima facie* moral wrongness.

However, I now want to bring out that the proposed wrong-making feature does not exhaust the problematic features of discrimination. On my view, discrimination is (prima facie) wrong because it constitutes a harm (in a broad sense) to the discriminate – but the problem does not stay there. The harm is not just any kind of arbitrary harm; it is harm due to the discriminatee's (believed) group membership: the socially salient property P. This is what makes discrimination especially pernicious; and this should be brought out and analysed by a theory of discrimination. I briefly sketch this idea in the next section, before I turn to applying my theory of discrimination to future generations.

[28] It should be noted that my account is similar to Scanlon's; see (Scanlon 1998).

[29] Note that the suggested wrong-making feature can capture the wrongness of even structural forms of discrimination. It focuses on the discriminatee, Y (who might be an individual or a group, supposing that we can make sense of groups being made worse off or treated as inferior). It can thus get a grip even on cases where there is no (individual or collective) discriminating agent, X, but rather a social entity involving social structures (which is something that, e.g., mental state accounts of the wrongness of discrimination would struggle with).

2.3. Injustice

Until now, I have not considered conditions (i) and (iv) of the above definition of discrimination:

> (i) there is a property, P, such that Y has P (or is believed to have P), and
> (iv) P is the property of being a member of a socially salient group.

To know where and when these conditions apply, we need to know what 'socially salient group' means. A standard way to understand the expression is this:

> "A group is socially salient if [and only if] perceived membership of it is important to the structure of social interactions across a wide range of social contexts."[30]

This seems to be on the right track, but we should dig deeper at this point. Why is social salience a relevant feature? Why should the definition of discrimination make reference to properties that are "important to the structure of social interactions across a wide range of social contexts"?

I propose that such a structure of social interactions becomes relevant for a definition of discrimination when it reflects systematic and unjustified social inequalities between groups of people – the kinds of inequalities which theories of social justice (and injustice) help us identify.[31] What makes discrimination especially problematic is that the harm done to discriminatees latches onto, reproduces, and over time exacerbates such social inequalities. Specifically, we take issue with instances of such harms when they are done to individuals or groups who are already on the systematically disadvantaged side of the social inequalities in question – and where this is, moreover, *due to* them belonging to this side in the first place.

To capture this, I propose the following analysis of 'socially salient group':

> A group is socially salient if and only if its members are subject to systematic, unjustified social disadvantages within the given social context.[32]

---

[30] (Lippert-Rasmussen 2014, 30).

[31] E.g., (Rawls 1999).

[32] I here want to note (although I do not have space to develop and defend) the upshot that this limits discrimination to socially disadvantaged groups. Those on the systematically advantaged side of the social inequalities in question can thus not be discriminated against, according to my account. This may seem counterintuitive at first, but arguably has the advantage of classifying, e.g., affirmative action (of the kind just recently ruled out as unconstitutional by the US Supreme Court) as non-discrimination. (Cf. https://www.washingtonpost.com/politics/2023/06/29/affirmative-action-supreme-court-

This analysis allows us to spell out a deeper problematic feature of discrimination. It is not only morally wrong but also unjust:

> *Injustice*: An instance of (group) discrimination is unjust when and because it reproduces and exacerbates systematic, unjustified social disadvantages.

Given this analysis, we can now plug in our preferred theory of social justice into the theory of discrimination (just as we did with our preferred account of moral wrongness – here: the broad harm account). Such a theory is needed to spell out, e.g., the relevant distributive patterns and the relevant currency of justice. It can then serve to provide us with the criteria to find whatever are the specific *grounds* of discrimination in a given context. In many of today's societies, these grounds will turn out to be the well-known categories of race, gender, disability, religion, and so on. And such exact categories arguably need to be specified, e.g., in legal statutes and institutional policies, in order to allow for the efficient and orderly application of such rules in real-life situations. However, behind any such specific categories, there should be a general criterion.[33] And this should be made explicit by a theory of discrimination – not least so that we, from time to time, can re-evaluate the specific grounds of discrimination that currently happen to be codified in our legal statutes and institutional policies.

The theory can now be summarised as follows:

> *Definition*: An agent, X, (group) discriminates against someone, Y, in context C by ϕ-ing if and only if:
>
> (i) there is a property, P, such that Y has P (or X believes that Y has P),
>
> (ii) by ϕ-ing, X treats Y worse than X would have treated Y, had Y not had P (or had X not believed Y to have P),
>
> (iii) it is because Y has P that X treats Y worse by ϕ-ing (or because X believes that Y has P), and
>
> (iv) P is the property of being a member of a socially salient group in C, where a group is socially salient in C if and only if its members are subject to systematic, unjustified social disadvantages within C.

ruling/).

[33] Such a criterion need not, of course, solely pick out single-dimensional categories. This approach may thus provide a new pathway for taking on the intersectionality challenge (a possibility that I'm hoping to explore further).

> *Moral wrongness*: An instance of (group) discrimination, ϕ-ing, by X against Y, on grounds of P, is (prima facie) wrong because X by ϕ-ing treats Y worse, in the following sense:
>
> (a) making Y worse off, or
>
> (b) treating Y as inferior (in the sense of demeaning),
>
> than X would have, had Y not had P (or had Y not been believed to have P).
>
> *Injustice*: An instance of (group) discrimination, ϕ-ing, in C, is unjust when and because it reproduces and exacerbates systematic, unjustified social disadvantages in C.

On this theory, then, group discrimination is a bridge concept. As such, it is a distinctive and particularly useful concept exactly because it connects wrongness to injustice: the moral domain (where individuals can be harmed, and thus *prima facie* wronged) to the political domain (where there are unjustified inequalities between social groups, i.e., social injustice). Normatively speaking, this theory brings out that discrimination is, in fact, doubly problematic. It is a driver of social injustice, and it operates by wronging (harming) already disadvantaged individuals – adding, as it were, injustice to injury.

3. Applying the theory to the problem of differential treatment of future generations

Is there conceptual space for wrongful and unjust discrimination against future generations in this framework? From the outset, it may seem so. I.e., it seems we could easily fit different cases of differential treatment of future generations into my taxonomy of four forms of discrimination, according to *Table 2*.[34]

[34] These descriptions are modeled on the Maastricht Principles' formulations; see (i–iii) in the first quote above.

	Differential treatment	Disparate impact
Intentional	(ii) The present generation (X) shifts the burden of responding to present crises (ϕ) to future generations (Y), by intentionally excluding their interests from political decision-making processes, because they are future generations (P). This amounts to treating these future generations worse. (iii) The present generation (X) accords less value (ϕ) to the lives and rights of future generations (Y) than the lives and rights of present generations, because they are future generations (P). This amounts to treating these future generations worse.	(i) The present generation (X) wastes, destroys, and unsustainably uses resources essential to human life, under the facially neutral rationale of "ensuring economic growth" (ϕ), while intending – or knowingly accepting – that the resulting environmental damages accumulate over time. This amounts to treating future generations (Y) worse.
Non-intentional	(iii*) The present generation (X) discounts (ϕ) the impacts and burdens of present conduct on the lives and rights of future generations (Y), not because they are future generations, but due to epistemic uncertainty concerning present actions' future consequences. This amounts to treating these future generations worse.	(i*) The present generation (X) wastes, destroys, and unsustainably uses resources essential to human life, under the facially neutral rationale of "ensuring economic growth" (ϕ), without being aware that the resulting environmental damages accumulate over time. This amounts to treating future generations (Y) worse.

Table 2: Four cases of differential treatment of future generations.

Does the attempt to fit these different cases of differential treatment of future generations into my taxonomy actually succeed? In the following three subsections, I will analyse the proposed cases in more detail to answer this question. I will examine, firstly, the proposed agent (X) and subject (Y) of discrimination; secondly, the proposed ground (P); and finally, the claim that the assessed actions (ϕ) amount to worse treatment.

3.1. The agent and subject of discrimination

Consider, first, the agent of discrimination (X). As the examples in *Table 1* (in section 2 above) demonstrated, it is conceivable and intuitively plausible that a discriminator may be a collective*:* e.g., a university (or other form of sufficiently organized collective entity with at least minimal collective intentionality[35]). Thus, even a generation, such as the present one – possibly partitioned into multiple collectives of citizens, represented by state officials who decide policies on their behalf – might be an agent of discrimination. Analogously, it is conceivable and intuitively plausible that the subject of discrimination (Y) may be a collective entity. Indeed, at a minimum, it may just be the group of people sharing a socially salient property, such as race or gender.

3.2. The socially salient property

Second, then, can 'being a future generation' be argued to be a socially salient property (P)? To answer this question, we first need to know how 'being a future generation' should be understood. As noted in the introduction above, cases of, e.g., unsustainable resource use by the current generation have been discussed under descriptions such as "discrimination by date of birth",[36] and "discrimination against future generations".[37]

"Date of birth", it should be noted here, is ambiguous as a ground of discrimination. It may refer to chronological age, such that the discrimination amounts to the (synchronic) disadvantageous treatment of certain age groups: e.g., a policy disadvantaging everyone under the age of 18 by not permitting them to vote. Or it may refer to birth cohorts, such that the discrimination amounts to the (diachronic) disadvantageous treatment of certain birth cohorts; e.g., a policy disadvantaging everyone born after a certain year, by raising the age threshold for pension benefits.[38] In the present context, it is the latter, the birth cohort reading, we are concerned with. However, this is still too general here, since we are not concerned with the disadvantageous treatment of different birth cohorts, across the board, but rather with the disadvantageous treatment of different birth cohorts marked by their different temporal status: future (as opposed to present or past), to be precise.

[35] Cf. (List et al. 2011). In an unpublished paper, I furthermore argue that there can be non-agential forms of discrimination, i.e., structural discrimination. It turns out that, properly described, cases of implicit bias discrimination are cases of such non-agential, structural discrimination: Berndt Rasmussen (mimeo) "Structural Dimensions of Discrimination".

[36] (Stern 2014).

[37] (Gardiner 2017).

[38] Note that both conceptions can be rendered equivalent by time-indexing chronological age: specifying age-at-t refers to a specific birth cohort (Gosseries 2015).

To make this clear, we should use 'future temporal status' as the relevant property P in this context. That is, we are here concerned with what may be called 'temporal status discrimination'.

So, can temporal status (past–present–future) be argued to be a socially salient property (P)? If we just think of a socially salient group as a group "perceived membership of [which] is important to the structure of social interactions across a wide range of social contexts",[39] this seems rather uncontroversial. Apart from epistemic challenges, lack of reciprocity, etc., in our interactions with the future, in any set of social contexts, there is the (nomological) *impossibility* for any generation to interact with past generations in ways that affect them; e.g., aiming to benefit or harm them (at least on the intuitively plausible assumption that there cannot be posthumous benefits and harms).

However, as proposed above, the key factor making the structure of social interactions relevant for discrimination is that it reflects systematic and unjustified social inequalities between groups of people – the kinds of inequalities which theories of social justice (and injustice) help us identify. Yet then it is rather doubtful that temporal status (past–present–future) qualifies as a socially salient property (P). This is so especially when we consider cases where it is the purported discrimination itself that gives rise to systematic and unjustified social inequalities between the present and future generations. And this seems to apply to the above cases of (i–iii), of resource depletion, of burden shifting, and of discounting of future lives. Absent already existing such inequalities, prior to the purported discrimination, the property of future temporal status does not yet qualify as 'socially salient'. And this, I want to emphasise, is as it should be: it captures the intuition that what makes discrimination especially problematic is that the (broad) harm done to discriminatees latches onto, reproduces, and over time exacerbates already existing social injustices. Thus, when it comes to the above cases (i–iii), they constitute the very choices that give rise to an injustice between the present and future generations – an injustice which then, on pain of circularity, cannot serve as a condition for classifying these cases as instances of intergenerational discrimination.

This result will turn out to be in line with the overall analysis of the differential treatment of future generations, once we consider the next hurdle for the discrimination account.

[39] (Lippert-Rasmussen 2014, 30).

3.3. The claim of "worse treatment"

The third question is whether the actions of the present generations in cases (i–iii) really amount to treating future generations worse. At first sight, this seems quite uncontroversial. Future generations will have to pay – with their money, health, security, opportunities, and even their lives – for the present generation's choices of resource depletion, burden shifting, and discounting of future lives. Surely, one might argue, they would have been better off, if the present generation had chosen resource conservation, had taken on the burdens themselves, and had valued future lives equally. Hence, by making these choices of (i–iii), the present generation treats future generations worse.

However, recall that on the above (broad) harm account of the wrongness of discrimination, 'treating worse' was interpreted in a person-affecting manner. Both the welfarist and non-welfarist components aim to capture what is intuitively problematic with discrimination: that people – individually or in groups – suffer from such differential treatment; that it is a wrong that is directed against them. Consider first the welfarist component, requiring that by ϕ-ing, X makes Y worse off, i.e., lowers Y's well-being compared to the counterfactual where Y had not had (or had not been believed to have) P. Now of course, as stated, the theory allows for discrimination to be directed against collectives of individuals. Still, the concern with well-being ties the harm account to the individuals constituting these collectives, rather than the collectives themselves. Claiming that a collective's well-being is lowered, such that the collective is harmed, presupposes that there are individual members whose well-being is lowered. Lowering the collective's well-being by, e.g., reducing the number of its individual members (such that the collective's total welfare is reduced), does not constitute a relevant harm: all else equal, no one suffers or is wronged by this.

Of course, such a person-affecting view is vulnerable to Parfit's infamous non-identity problem when it comes to future people.[40] The present generation's choice between, e.g., resource depletion or conservation will have global effects on people's consumption and traveling patterns, education and lifestyle choices, etc. It will thus arguably affect which people will be born to such an extent that already within the next few generations' time, there will be virtually no overlap between the sets of people who would exist under either alternative of the choice. But then, there will be no individual member of these future generations who is made worse off by the present generation's choice to deplete rather than conserve resources. Had the present generation chosen conservation, none of them would have existed. Hence, no one is treated worse in this sense. As long as all have a life worth living – and this

[40] (Parfit 1986, 302–31).

is of course an important proviso – arguably no one suffers or is wronged by the choice to deplete.

How about the other, non-welfarist sense of being treated worse: being treated as inferior? I suggested to fill this in by reference to Hellman's demeaning account: discrimination is wrong when it is demeaning, in the sense that it, first, "expresses that a person or group is of lower status [...] and, second, the actor or institution expressing this meaning must have sufficient social power for this expression to have force".[41]

The first condition may not be satisfied in all of the cases analysed above. E.g., if the present generation discounts the impacts and burdens of their choices on the lives and rights of future generations, *not* because they are future generations, but due to the epistemic uncertainty concerning present actions' consequences for these future generations (case iii'), this need not express any inferiority of status. Such a case will then not be an instance of wrongful, *qua* demeaning, discrimination. Still, some of the other cases explicitly mention a lowering of status: e.g., if the present generation accords less value to the lives and rights of future generations than their own, just because they are future generations (case ii), or when it intentionally excludes their interests from political decision-making processes, thereby arguably denying their political equality (case iii).

The remaining question is then whether Hellman's other condition is satisfied in these cases as well. Does the agent expressing this meaning have sufficient social power for this expression to have force? This is doubtful. Of course, in one sense, the present generation has a clear power overtake on future generations: it is in a position to call all the shots as to resource depletion, of burden shifting, and of discounting of future lives – choices that undoubtedly affect the living conditions for future generations. But the main issue for Hellman's account is one of *social* power, which presupposes what we may call a "community of social meaning" within which the present generation's status lowering has force. To assess whether this holds for our cases, we need to consider the question of who is the relevant audience generating the required uptake, such that the expression of the inferiority of future generations has force. Is it the future generations? If so, it seems doubtful that the present generation really would have the social power to mark future generations as inferior *in their own view*. Is it all generations, i.e., all of humanity? Again, it seems doubtful that *one* generation would have the social power to mark future generations as inferior in the view of all generations to come (and they definitely lack this power for past generations). Is it just the present generation, then? But then, if it is merely an internal status lowering that does not carry over to future

[41] (Hellman 2017, 102).

generations (more than that it affects what living conditions they will find themselves with) it no longer seems to concern any other generations, and thus it seems doubtful that it really treats *them* worse by demeaning them. (And note that the just mentioned living conditions that future generations will find themselves with fall under the welfarist component of the broad harm concept – which, again, leads us back to the non-identity problem.)

At the bottom of the problem from lack of "community of social meaning" lies the idea that treating someone worse by treating them as inferior, in the sense of demeaning, tracks a relational concern, which presupposes a community in which social meaning is expressed and given uptake. And it is doubtful that there is such a community between temporally non-adjacent generations. If this is so, Hellman's second condition is not fulfilled, and hence, the choices of present generations fail to be demeaning of future generations.

In sum, if these choices neither make individuals belonging to future generations worse off, nor demean them, we cannot classify them as instances of *wrongful* discrimination on the proposed broad harm account. Indeed, on the above definition, we cannot even classify them as *discrimination* in the first place: they do not constitute worse treatment, neither in the welfarist nor in the non-welfarist sense.

4. Conclusions and upshots

In this paper, I have set out to assess whether the present generation's choices – of resource depletion, burden shifting, and discounting of future lives – can be classified as discrimination: as disadvantageous treatment of future generations due to their future temporal status. Drawing on my preferred theory of discrimination, I have examined whether such choices can fit the taxonomical and normative framework of this theory. Such a fit would have repercussions on the moral and political-philosophical evaluation of these choices. It might, further, call for bolder advances in legal theory, to find new ways for litigation under existing bodies of discrimination law.

Despite a promising start, granting that both discriminator and discriminatee may indeed be collectives such as generations, I eventually arrived at the conclusion that these choices do not fall under my definition of discrimination. This is so since, firstly, the property of future temporal status does not qualify as 'socially salient' on my injustice-based criterion of social salience, which generates context-specific grounds of discrimination. And secondly, a closer analysis of the "worse treatment" claim made for each of the cases in *Table 2* revealed that this claim does not hold up. In the welfarist sense of making future generations worse off, it suffers from the infamous non-identity problem. And in the non-welfarist sense of treating them as

inferior in the sense of demeaning, it suffers from what we may call the problem from lack of "community of social meaning".

My analysis is admittedly still sketchy and open for improvements. It might, e.g., turn out – although I am highly doubtful of this – that there are independent reasons to favour a disrespect account of the wrongness of discrimination (as a pure mental state account).[42] On such an account, arguably, the present generation's choice to accord less value to the lives and rights of future generations than those of present generations is morally faulty. Plugging this account into my definition of discrimination would render a theory that could classify the present generation's choices as wrongful discrimination. Theoretical options, such as this one, surely deserve further scrutiny.

However, if my arguments hold up, there is something to be said for this negative result as well. For one thing, my theory of discrimination can explain why people sometimes are tempted to describe the present generation's choices to deplete resources, etc., as "discrimination by date of birth" or "discrimination of future generations". At first sight, future temporal status seems to be a socially salient category, and future generations seem to be harmed (in a welfarist or non-welfarist sense) by these choices. It takes a more careful analysis – employing a coherent theory of discrimination – to reveal that this is not so. Further, my analysis also helps explain why virtually no climate lawsuits have appealed to discrimination law when it comes to climate change affecting future generations in the Maastricht Principles' sense: as "those generations that do not yet exist but will exist".[43] Such a legal move may simply not be the most feasible route for combating climate change, given the conceptual and normative difficulties it carries.

On the other hand, and more practically speaking, it may seem worrying that my negative result might dissuade climate activists' attempts to make use of discrimination laws in climate lawsuits, when this *is* feasible. E.g., Axel Gosseries has examined the viability of litigation strategies that appeal to age-discrimination statutes and policies, under EU, US or domestic laws;[44] a proposal that has recently been further developed by Refia Kaya.[45] Concerning this worry, and these constructive suggestions, I want to end on a pragmatic note. In order to solve or mitigate the ongoing climate crisis, we should use whatever reasonable means available. Discrimination law is, of course, distinct from the philosophical analysis employed here. In legal contexts where such laws currently are formulated in a way that it makes climate litigation possible, we should go for it – without any philosophical headache.

[42] (Eidelson 2015).

[43] https://www.rightsoffuturegenerations.org/the-principles.

[44] (Gosseries 2015).

[45] (Kaya 2019).

References

Alesina, Alberto, Michela Carlana, Eliana La Ferrara, and Paolo Pinotti. 2018. 'Revealing Stereotypes: Evidence from Immigrants in Schools'. Working Paper 25333. National Bureau of Economic Research. https://doi.org/10.3386/w25333

Altman, Andrew. 2016. 'Discrimination'. In The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta, Winter 2016. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/discrimination/

Arneson, Richard. 2006. 'What Is Wrongful Discrimination?' San Diego Law Review 43 (4): 775. https://digital.sandiego.edu/sdlr/vol43/iss4/4

Attfield, Robin. 2010. 'Global Warming, Equity and Future Generations'. Human Ecology Review 17 (2).

Beeghly, Erin. 2017. 'Discrimination and Disrespect'. In The Routledge Handbook of the Ethics of Discrimination. Routledge.

Berndt Rasmussen, Katharina. 2019. 'Harm and Discrimination'. Ethical Theory and Moral Practice 22 (4): 873–91. https://doi.org/10.1007/s10677-018-9908-4

———. 2020. 'Implicit Bias and Discrimination'. Theoria 86 (6): 727–48. https://doi.org/10.1111/theo.12227.

———. 2023. 'Age Discrimination: Is It Special? Is It Wrong?' In Ageing without Ageism?: Conceptual Puzzles and Policy Proposals, edited by Greg Bognar and Axel Gosseries, 0. Oxford University Press. https://doi.org/10.1093/oso/9780192894090.003.0002

Campbell, Terry Ann. 2019. 'Environmental Constitutionalism: Marrying the Due Process Clause and the Equal Protection Clause with Climate Change'. SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=3457043

Eidelson, Benjamin. 2015. Discrimination and Disrespect. Oxford University Press. http://www.oxfordscholarship.com.ezp.sub.su.se/view/10.1093/acprof:oso/9780198732877.001.0001/acprof-9780198732877

Franco, Ana María Suárez, and Sandra Liebenberg. 2023. 'The Maastricht Principles on the Human Rights of Future Generations'. Future Generations. Global Policy Forum. https://www.globalpolicy.org/en/multilateralism/future-generations

Gardiner, Stephen M. 2017. 'Geoengineering: Ethical Questions for Deliberate Climate Manipulators'. In The Oxford Handbook of Environmental Ethics, edited by Stephen M. Gardiner and Allen Thompson, 0. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199941339.013.44

Gardner, Jim. 1978. 'Discrimination against Future Generations: The Possibility of Constitutional Limitation'. Environmental Law 9 (1): 29–60. https://heinonline.org/HOL/P?h=hein.journals/envlnw9&i=37

Gosseries, Axel. 2015. 'Environmental Degradation as Age Discrimination'. E-Publica 2 (2): 26–39.

Hellman, Deborah. 2008. When Is Discrimination Wrong? Cambridge, MA: Harvard University Press.

———. 2017. 'Discrimination and Social Meaning'. In The Routledge Handbook of the Ethics of Discrimination, edited by Kasper Lippert-Rasmussen. London; New York: Routledge.

Kaya, Refia. 2019. 'Environmental Vulnerability, Age and the Promises of Anti-Age Discrimination Law'. Review of European, Comparative & International Environmental Law 28 (2): 162–74. https://doi.org/10.1111/reel.12279

Khaitan, Tarunabh. 2015. A Theory of Discrimination Law. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199656967.001.0001/acprof-9780199656967

Lavergne, Gary M. 2010. Before Brown: Heman Marion Sweatt, Thurgood Marshall, and the Long Road to Justice. University of Texas Press.

Lippert-Rasmussen, Kasper. 2014. Born Free and Equal?: A Philosophical Inquiry into the Nature of Discrimination. Oxford; New York: Oxford University Press.

———. 2017. The Routledge Handbook of the Ethics of Discrimination. Routledge Handbooks in Applied Ethics. New York: Routledge. https://ezp.sub.su.se/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1581310&site=ehost-live&scope=site

List, Christian, Philip Pettit, Christian List, and Philip Pettit. 2011. Group Agency: The Possibility, Design, and Status of Corporate Agents. Oxford, New York: Oxford University Press.

Mendoza, José Jorge. 2017. 'Discrimination and Immigration'. In The Routledge Handbook of the Ethics of Discrimination, edited by Kasper Lippert-Rasmussen. London ; New York: Routledge.

Moreau, Sophia. 2010. 'What Is Discrimination?' Philosophy and Public Affairs 38 (2): 143–79.

———. 2013. 'In Defense of a Liberty-Based Account of Discrimination'. In Philosophical Foundations of Discrimination Law, edited by Deborah Hellman and Sophia Moreau, 71–86. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199664313.001.0001/acprof-9780199664313-chapter-5

———. 2017. 'Discrimination and Freedom'. In The Routledge Handbook of the Ethics of Discrimination, edited by Kasper Lippert-Rasmussen. London ; New York: Routledge.

Nguyen, Ylam. 2017. 'Constitutional Protection for Future Generations from Climate Change'. UC Law Environmental Journal 23 (1): 183. https://repository.uclawsf.edu/hastings_environmental_law_journal/vol23/iss1/2

Parfit, Derek. 1986. Reasons and Persons. Oxford, New York: Oxford University Press.

Rawls, John. 1999. A Theory of Justice: Revised Edition. Cambridge, MA: Belknap Press.

Scanlon, Thomas. 1998. What We Owe to Each Other. Cambridge, Mass.: Belknap Press of Harvard University Press.

Stern, Nicholas. 2014. 'ETHICS, EQUITY AND THE ECONOMICS OF CLIMATE CHANGE PAPER 2: ECONOMICS AND POLITICS'. Economics & Philosophy 30 (3): 445–501. https://doi.org/10.1017/S0266267114000303

Sulyok, Katalin. 2023. 'A rule of law revolution in future generations' litigation: intergenerational equity and the rule of law in the Anthropocene'. https://doi.org/10.25360/01-2023-00005

Zschirnt, Eva, and Didier Ruedin. 2016. 'Ethnic Discrimination in Hiring Decisions: A Meta-Analysis of Correspondence Tests 1990–2015'. Journal of Ethnic and Migration Studies 42 (7): 1115–34. https://doi.org/10.1080/1369183X.2015.1133279

Krister Bykvist[1]

Escaping the Impossibility Theorems in Population Ethics[2]

Decision-makers are in a hurry to find morally justified responses to climate change. Population ethicists have thrown a spanner in the works by formulating various impossibility theorems that show that no theory about the value of population change can satisfy all the conditions we think such a theory must satisfy. What shall we do, if we do not know which condition(s) to give up? One relatively unexplored option is to view the satisfaction of a condition as a matter of degree, as Geoff Brennan recently has suggested (in the context of Arrow's impossibility theorem). This opens up the possibility that some theories might overall come closer to full satisfaction of the conditions than others. In my paper, I shall explore various versions of this idea and see how far they will take us. In particular, I will make use of the famous Kemeny-measure of distance and show that this will rule out all population theories that are indifferent between some of the alternative populations in the Mere Addition Paradox. I will also discuss factors beyond distance that are relevant for theory choice.

[1] Dept. of Philosophy, Stockholm University, Institute for Futures Studies, Stockholm

[2] For very helpful feedback, I would like to thank the audience at the internal work-in-progress workshop of the Climate Ethics program held in Spring 2023, especially Gustaf Arrhenius, Tim Campbell, Hilary Greaves, Daniel Ramöller, and Joe Roussos.

1. Introduction

Decision-makers are in a hurry to find morally justified responses to climate change and other urgent issues that involve decisions that will have effects on future populations. But the population ethicists have not been especially helpful. We have thrown a spanner in the works by formulating various impossibility theorems that show that there is no acceptable reaction to climate change if we take into account the value of population change. More precisely, these theorems show that there is no theory about the value of population change that satisfies a set of very plausible conditions we are inclined to think a theory should satisfy.[3] Given that no theory can satisfy all of these conditions, what shall we do?

The main options are to

(1) 'put your hands up in the air': utter despair and moral paralysis, for population ethics is doomed to be inconsistent.[4]

(2) 'not care': argue that it is a mistake to think the impossibility theorems in population ethics are relevant for moral justification;

(3) 'drop a condition': sit down, do some serious philosophical reflection, and try (again) to work out which condition(s) to drop;

(4) 'hedge': keep all the different theories on the table, assign credences to them, compare the values the theories assign to populations, and, in analogy with what we should do under empirical uncertainty, apply some suitable decision-theoretic principle.[5]

(5) 'think that a miss is *not* as good as a mile': instead of just judging whether a theory satisfies or fails to satisfy a certain condition, we can see whether it gets *closer* or further *away* from satisfying the condition, as Geoff Brennan recently has suggested in the context of Arrow's impossibility theorem (Brennan 2015, see also Brennan and Braurmann 2006). Moreover, even if no theory can satisfy all of the conditions, some might come overall *closer* to satisfying them than others. Hopefully, we could seek guidance from the theories that rank higher.

In this paper, I shall explore various versions of the closeness approach and see how far they will take us. To say that one will *explore* something is a philosopher's jargon

[3] See, for instance, one of the leading spanner throwers Arrhenius (forthcoming). For an impossibility theorem in a probabilistic setting, see Arrhenius ibid.

[4] That population ethics is inconsistent is seriously considered in Arrhenius (forthcoming).

[5] For an example of this approach, see Bykvist (2022) and Ord & Greaves (2017).

for saying that one has not yet made up one's mind about the issues, or failed to reach a conclusion. As you will see, I am not sure that the closeness idea can take us far enough. I have excuses for this undecidedness. The issue I am going to discuss involves a lot of uncharted terrain, and the issues are very complex. But I hope this exercise in 'axiological escapology', as we may call it, still can teach us something important, and that it is not just a failed escape act from the chains of the impossibility theorems.

Before I explore the closeness approach, I will introduce the impossibility theorems in population ethics (one simple version, there are many others!), and say a few words about the other alternative reactions to impossibility theorems and why it is worth exploring the closeness approach.

2. Impossibility theorems in population ethics

In general, to show an impossibility theorem is to collect a set of intuitively plausible conditions on a certain kind of theory and prove that they are logically inconsistent. This is what Arrow did for theories of social choice (with interpersonally incomparable ordinal preferences), and this is also what is done in population ethics for theories about how we should value populations. Now, there are many different impossibility theorems in population ethics. Here I will only present a very simple version, since it is easier to work with; it can be discussed informally without getting into technical details. It should be noted that the conditions of this version are not as plausible as the ones of the more complex formal ones.[6]

The conditions for what is often called the Mere Addition Paradox can be stated informally as follows.

> *Mere Addition*, a population that differs from another only in that it contains some extra lives all worth living is *at least as good* as the smaller population.
>
> *Non-Anti Egalitarianism (NAE)*: a same-sized population with both greater total and average wellbeing, distributed perfectly equally, is better.
>
> *Avoidance of the Repugnant Conclusion (Avoidance of RC)*: a vast population with lives barely worth living is worse than a much smaller population with lives of very high wellbeing.[7]

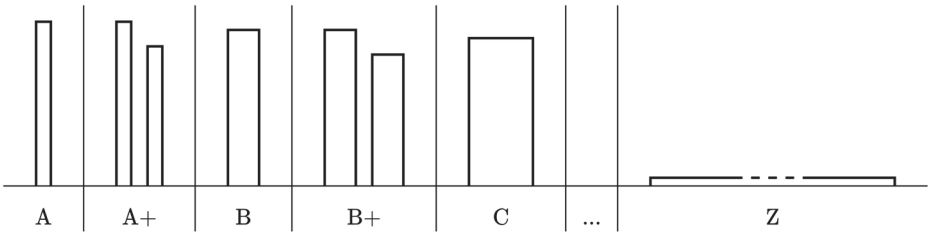
[6] For more plausible and formally developed theorems, see Arrhenius (forthcoming) and Blackorby et al. (2005).

[7] Strictly speaking, a theory avoids RC when it states that a vast population with lives barely worth living is not better than a much smaller population with lives of very high wellbeing. But the stronger

The fudge words 'vast', 'much smaller', 'barely worth living', and 'very high' can be avoided in the more formally precise statements of the conditions.

Here is an illustration of the impossibility of satisfying all the conditions above.

Fig. 1



A is a population with lives of very high wellbeing. By Mere Addition: A+ is at least as good as A. By NAE, B is better than A+. This implies that B is as least as good as A, by transitivity of at least as good as. Now repeat this for B, B+, and C and so on until you reach Z, a vast population with lives barely worth living, and we can conclude that Z is at least as good as A. But this contradicts the claim, stated by Avoidance of RC, that A is better than Z.

Strictly speaking, the conditions that generate the impossibility should include these background conditions:

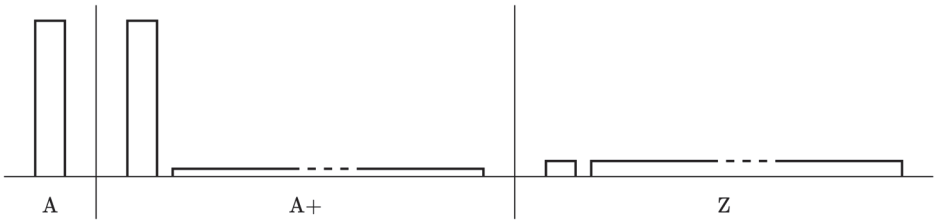
Transitivity: of at least as good as;
Measurability: assumptions about the structure and measurability of wellbeing that make it possible to construct a sequence as the one above and to talk about total, average, and equal wellbeing;
and
Domain Richness: there are possible populations like A, A+, B, B+, ..., and Z.[8]

A much shorter version of the paradox is this.

condition in the text is usually what motivates people to demand that a theory avoids RC.

[8] The strong measurability assumptions can be avoided. See Arrhenius (forthcoming).

Fig. 2



However, one could object here that the move from A to A+ is very problematic because it creates an enormous amount of inequality, or that the move from A+ to Z implausibly requires us to pull down the better off people to the level of barely worth living. However, each move in the long version can be justified by invoking weaker principles than the ones listed above (and relaxing the measurability assumption). But I will work with the short version to avoid unnecessary complexity.

Another simplifying assumption is that I shall work with a stronger version of Mere Addition, according to which population A+ is always *better* than population A. Furthermore, I shall put aside the conditions *Transitivity*, *Measurability*, and *Domain Richness*. I shall also assume that the theories satisfy full comparability (i.e., there are no gaps in the ordering). Finally, I shall assume that the set of alternatives to be compared (the set of populations) are finite, as well as the set of people in each population. I will come back to some of these simplifications later.

3. Reactions to the impossibility theorem

Of the listed options, I think we should not 'throw up our hands in the air' unless we have shown that none of the others work.

The option of not caring is worth considering, since some might argue that there is something deeply mistaken about the framework. Why should we think that the value of populations matters, and why think that it is a function of the wellbeing of its members? This smacks of old-fashioned utilitarianism. Why should non-utilitarians care about this? However, this concern isn't exclusive to utilitarians; everyone should care about how our current actions affect the wellbeing of future people. One should not be completely indifferent to the possibility of making future lives miserable, or barely worth living. Of course, this is not to deny that many other things matter as well, such as our rights and duties to contemporaries, but one factor to take into account is the wellbeing of future people. If all else is equal, this is the factor that determines what we should do.

The idea that only individual wellbeing matters for the value of a population can

also be relaxed. For example, the value of a life can be in part determined by moral, artistic, and athletic achievements. Finally, we can even state the impossibility theorems directly in terms of reasons or obligations to bring about various population changes, without invoking the value of populations or the value of outcomes.[9] But, again, for simplicity, I will stick to the wellbeing framework in this paper.

The ideal option is of course to drop a condition, if we know which one to drop. The problem is that we do not know, or many of us do not. Despite extensive philosophical reflection on these issues, there is still wide-spread disagreement and undecidedness among population ethicists. Since climate change requires immediate action, we need some guidance on what to do now, even though we are undecided about which condition to drop.

Hedging could be an option here (which I have explored in Bykvist (2022)). The idea is to view the choice situation like this, where numbers represent the value of populations A, A+, and Z according to some theories, T1, T2, and T3:

Fig. 3

Alternatives	**T1** p1	**T2** p2	**T3** p3
A	1	1	1
A+	2	-1	2
Z	3	0	0

Each theory considered satisfies two out of the three conditions. T1 satisfies Mere Addition and NAE, but not Avoidance of RC. T2 satisfies NAE and Avoidance of RC, but not Mere Addition. T3 satisfies Mere Addition and Avoidance of RC, but not NAE.

To decide which population to bring about, we need to somehow weigh the probabilities (credences for the different theories) p1, p2, and p3, against the values assigned to the populations by the theories T1, T2, and T3. One major challenge for this approach is to show that it makes sense to compare values across *different* theories. While I think it does make sense in some cases, this is controversial.[10] Thus, it is worthwhile to explore the last option.

[9] For a deontic impossibility theorem that only invokes 'ought' and 'permissible', see Arrhenius (2021).

[10] For a critical discussion of some existing proposals and a defence of a new one, see MacAskill et al (2020) and Riedener (2021).

4. Satisfying a condition is not an *all or nothing* affair

The guiding idea of this approach is that instead of just judging that a theory satisfies or fails to satisfy a certain condition, we can say that a theory gets closer or further away from satisfying the condition (for short, 'closer or further away from the condition'). Moreover, even if no theory can satisfy all of the conditions, some might come overall closer to satisfying them than others. This degree of closeness can be understood in different ways, but a plausible closeness account must validate:

> *Closeness Dominance*
> If, for every condition C, T1 is at least as close to C as T2 is, and for some condition C', T1 is closer to C' than T2 is, then T1 is closer overall to satisfying the conditions than T2 is.
>
> *Equal Closeness*
> If, for every condition C, T1 is exactly as close to C as T2 is, then T1 is exactly as close to all conditions as T2 is.
>
> *Inclusion*
> If T1's C-violations are a proper subset of T2's C-violations, then T1 is closer to C than T2 is.
>
> *Identity*
> If T1's C-violations are exactly the same as T2's C-violations, then T1 is exactly as close to C as T2 is.

While these principles have some applicability, but the first two require closeness comparisons between different theories regarding a certain condition. None of them requires closeness comparisons across conditions, i.e., that one theory is closer to a certain condition than another theory is to another condition. But this also shows its limitations. Ideally, we would like to make overall closeness comparisons when theories differ in how close they are to a whole set of conditions.

I shall consider three accounts of closeness: a value-based approach (defended by Brennan (2015)), a proportion-based approach, and a ranking-distance approach —which is the one I will end up favouring if combined with a proportion-based approach. To simplify the discussion, I will assume that all conditions have the same weight. This is unrealistic, since we might have more confidence in some conditions than in others. In section 7, I will briefly discuss the significance of dropping this idealization.

5. Value-based approach

To explain the motivation behind his value-based approach, Brennan usefully invokes an analogy with bananas and apples. Suppose you wish to eat 10 bananas and 7 apples a week, but you can't afford this fruit consumption. You should not declare yourself an *apple person* or a *banana person* and forget about the other fruit. You should trade off the fruits so that you get an ideal combination of apples and bananas, which normally means that you will give up some of both. Brennan suggests something similar for impossibility theorems. When you realize that not all conditions can be jointly satisfied by a theory, you should not just go for some conditions and forget about the others. You should trade off some conditions against others until you find a theory that is best in terms of an ideal trade-off between the different conditions.

How does Brennan's approach work more exactly? First, we need to identify for each condition 'the underlying value' that this condition is 'supposed to promote' (Brennan 2015). Then, we form a metric that 'reflects the degree to which a procedure fails' to meet the condition. A theory fails to meet a condition when the theory promotes the underlying value below a certain *threshold* level.

This suggests that a theory's closeness to a condition is the difference between the amount of value 'promoted' by the theory and the threshold of value set by the condition. A theory's closeness to the set of conditions is then some strictly decreasing function of all the value differences between the theory and the conditions.

Since Brennan talks about Arrow's Impossibility Theorem and Sen's Liberal Paradox, the conditions he has in mind are: *Universal Domain*, *Independence of Irrelevant Alternatives*, *Pareto*, *Non-Dictatorship*, *Transitivity of 'at least as good as'*, and *Minimal Liberty*. Brennan concedes that developing a metric for each of these conditions is a great challenge. But he suggests that for some of the conditions it is pretty easy. For example, he claims that we can measure how well a theory does in terms of Universal Domain by the proportion of possible individual rankings that have to be ruled out. Furthermore, when he considers the Pareto-principle, he suggests that the value it promotes is preference satisfaction and that the metric should be defined in terms of *distance from a Pareto-optimal frontier* (the set of Pareto-optimal social states).

This measure should, with suitable constraints on the underlying values, be able to satisfy the general principles: Closeness Dominance, Closeness Equality, Inclusion, Identity. But there are some problems with the account, especially if we want to apply it to the Mere Addition Paradox.

First, it seems very questionable that each of these conditions has a *unique* underlying value that is supposed to be promoted. Putting aside *Transitivity* and

Domain Richness, which might be exceptions, what is the *unique* promotion-worthy value underlying the *Mere Addition Principle*, the *Avoidance of RC*, and *NAE*, respectively? Each of these conditions can be accepted for a variety of reasons, and from very different evaluative standpoints. That is especially clear for *NAE*, which can be accepted by pure egalitarians, total utilitarians, average utilitarians, and leximiners. But it is also clear that the Mere Addition Principle can be accepted by total utilitarians and person-affecting views, and the *Avoidance of RC* can be accepted by average utilitarians, critical level utilitarians, leximiners, and various perfectionist theories. Indeed, that a condition can be accepted by very different moral outlooks is one of the main reasons why we assume it in the first place, since a condition that only a few outlooks would accept can more easily be rejected.

Second, even if we assume that there are values underlying each condition, why assume that there is a threshold for each of these values? And if there is threshold, how do we decide where it is?

Finally, and more importantly, Brennan asks us to assess theories according to how well they trade off the values underlying the conditions. But this is odd, since the conditions were supposed to constrain value trade-offs. For example, to accept the Avoidance of RC is to accept that no number of barely worth living people can together be more valuable than a smaller number of very well-off people. So, when Brennan asks us to judge theories according to how well they trade off various values, we seem to be back to where we started. We have a set of values and we want to know how to aggregate them. For instance, we want to know how to weigh quality of wellbeing against quantity of wellbeing. The impossibility theorems were generated by listing all plausible conditions on such trade-offs. Unless Brennan can show us which condition(s) to drop, we have not escaped the impossibility theorems.

6. The proportion-based approach

According to this approach, a theory's closeness to a condition C is identified with the proportion of its C-violations. The greater proportion of C-violations a theory has, the further away the theory is from condition C. (The account could of course be restated in terms of proportions of satisfactions of a condition.)[11]

Overall closeness is then some strictly decreasing function of the closeness measures for each condition. For example, if we can measure the exact proportion of violations for each condition, we can average these measures to get the overall closeness to the set of all conditions.

[11] A similar approach has been defended by Campbell and Kelly (1994), who construct a measure of degrees of Non-Dictatorship satisfaction in terms of the percentage of total alternatives someone has dictatorial power over.

I think this account is on to something, for proportions of violations seem to be a relevant factor for closeness. But it cannot be the whole story, for not all violations are *on a par*. If a condition states that X-alternatives are *better* than Y-alternatives, then a theory that says that X-alternatives are *equal in value* to Y-alternatives seems closer to the condition than a theory that says that X-alternatives are *worse* than Y-alternatives. This suggests the following general principle:

> If T1 swaps C's ranking of the alternatives and T2 ties the alternatives, then T2 is closer to C than T1 is, other things being equal.

In short, swaps take a theory further away than ties, other things being equal.

Consider the Avoidance of RC, and the A- and Z-populations from above. A theory that states that A-populations are equally as good as Z-populations is closer to Avoidance of RC than a theory that says that A-populations are worse than Z-populations. So, even if two theories can have the same proportion of C-violations, one can come closer to C than the other because its violations are ties rather than swaps.

Of course, this does not disqualify the proportion-account, if we understand it as saying that the proportion of violations matter, when *other things are equal*:

> If T1's proportion of C-violations are greater than T2's, then T1 is closer to C than T2 is, *other things being equal*.

7. Ranking-distance approach

The ranking-distance approach defines a theory's closeness to a condition in terms of the *distance* between the theory's ranking and the ranking(s) given by a condition. There are three notions that need to be explained here: a theory, the notion of a distance, and the notion of the ranking(s) given by a condition. For simplicity, I will work with a course-grained notion of a theory, according to which a theory is just an ordinal ranking of populations. A more fine-grained notion of a theory would include an *explanation* of why a given ranking of states of affairs holds.

The notion of distance I am going to work with is the popular Kemeny-metric (Kemeny 1959), which has been used in the contexts of information technology and social choice. It has the virtue of being very simple, capturing some intuitive ideas about distance and closeness, being impartial between conditions, and not requiring anything more than an ordinal ranking of alternatives. (I aim to deal with alternative metrics in the future.)

The intuitive idea is that the distance between two rankings is the number of

minimal changes one has to apply in order to get from one ranking to the other. To define it more precisely, note first that a ranking R can be represented as a set of ordered pairs of alternatives, such that a pair (x, y) belongs to R if and only if R ranks x at least as highly as y. Now, the distance between two rankings, R1 and R2, is simply the number of ordered pairs that belong to either R1 or R2 but not to both of these rankings. Finally, the *total* distance between a ranking R and a *set* of rankings is the *sum* of distances between R and each ranking in the set.

Consider the following example of distances between individual rankings (i.e., all the minimal moves required to swap the top-ranked and the bottom-ranked alternatives).

Fig. 4

x		x		x		x, z		z		z		z		
y	=>	y, z	=>	z	=>	y	=>	x	=>	x, y	=>	y		
z				y				y				x		
	1		1		1		1		1		1			

Here, the distance between each adjacent pair of rankings is 1. The distance between the first and the last rankings is 6.

The notion of the rankings given by a condition is more difficult to spell out. One option is to think about the rankings given by a condition C as *all* the possible complete rankings that satisfy C. The distance between the theory and C is then the total distance between the theory and the set of all the C-complying rankings. This is a non-starter, however. On this account, no theory can be at zero distance to a condition (thus, no theory is maximally close to a condition), since any theory is at a non-zero distance to some of the C-complying rankings.[12] But we know that some theories do satisfy and thus come maximally close to some of the conditions.

Another option is to take all the rankings that satisfy condition C and then identify the ranking(s) that minimizes the distance to all other C-satisfying rankings. Call these the *representative C-ranking(s)*. The closeness of a theory to a condition C is defined by the distance between the theory and the representative C-ranking(s).

On this account, a theory can be maximally close to a condition, but the obvious problem is that this holds only if it is *identical* to the representative C-ranking.[13] Any

[12] This holds for all plausible measures of a ranking's total distance to a set of rankings, for only a ranking that is identical to all rankings in the set is overall maximally close to the set.

[13] This holds for all measures of distance, for only identical rankings are maximally close to each other on any adequate distance measure.

other theory is at a non-zero distance from the condition. So, the account will violate the trivial constraint that if two different theories both satisfy a condition C, then they are both maximally close to satisfying it, i.e., they both have the distance value 0 with regard to C.

A better idea is to look only at the relevant *sub-rankings* in theories. If a condition ranks x versus y (given that x and y stand in the appropriate relation), then we only look at how theories rank x versus y. This solves the problems with the previous accounts, since if the condition ranks x over y, then *any* theory that ranks x over y is maximally close to the condition given that the distance is 0.

This account also validates the principle that a swap takes us further away than a tie comes out as obviously true.[14] x = y is always closer to x > y than y > x is. The distance between x = y and x > y is 1, and the distance between x < y and x > y is 2, as the following diagram shows.

Fig. 5

x				y
y	=>	x, y	=>	x
	1		1	

Now the conditions we are discussing do not just rank two populations; they rank *any pair* of populations that stand in certain relations to each other, spelled out by the relevant condition: in any (A, A+)-pair, the A+-population is better than the A-population; in any (Z, A+)-pair, the Z-population is better than the A+-population; in any (A, Z)-pair, the A-population is better than the Z-population. So, in order to decide how close a theory is to a certain condition C it is not enough to look at how close a theory comes to C's ranking of a certain pair; we need to look at how close it comes to the C's ranking of *each* pair, or each *C-ranking,* as we may call them. More precisely, to see how close a theory T is to C, the idea is to first look at how close the theory comes to each C-ranking. The distance between T and C is then the sum of the distances between T and each C-ranking. In order to see how close a theory is to *all* conditions, we should sum the distances between the theory and each condition.

[14] Other distance measures validate this too. For example, the Duddy-Piggins measure (Duddy & Piggins 2012) and the Cook-Seiford measure (Cook & Seiford 1978). Note, however, that not all distance measures will validate this. For example, the Hamming distance measure (Hamming 1950) will not validate it, since it defines the difference between two rankings as the number of (unordered) pairs of objects for which the rankings disagree. This means that the Hamming distance between the first and the second ranking is 1 and so is also the distance between the first and the third ranking.

The picture is this. Assume that we have three theories, T1, T2, and T3, which provide the following rankings.

Fig. 6

T1	T2	T3
A1+ > A1 A2+ > A2 ⋮ Ak+ > Ak	A1+ = A1 A2+ > A2 ⋮ Ak+ > Ak	A1+ < A1 A2+ < A2 ⋮ Ak+ < Ak
Z1 < A1+ Z2 < A2+ ⋮ Zl < Al+	Z1 < A1+ Z2 > A2+ ⋮ Zl > Al+	Z1 = A1+ Z2 = A2+ ⋮ Zl = Al+
A1 > Z1 A2 > Z2 ⋮ Am > Zm	A1 > Z1 A2 < Z2 ⋮ Am < Zm	A1 > Z1 A2 > Z2 ⋮ Am > Zm

Note that T2 illustrates the possibility that how well a theory fares with respect to a condition can vary from one case to another (e.g., A1+ = A1, but Ai+ > Ai, for all other i). These theories will show the following closeness distances to MA, NAE, and Avoidance of RC.

Fig. 7

	T1	T2	T3
Mere Addition (MA)			
A1+ > A1	0	1	2
A2+ > A2	0	0	2
⋮	⋮	⋮	⋮
Ak+ > Ak	0	0	2
Distance to MA	a1 = 0 + 0 + …+ 0	b1 = 1 + 0 +…+ 0	c1 = 2 + 2 +…+ 2
NAE			
Z1 > A1+	2	2	1
Z2 > A2+	2	0	1
⋮	⋮	⋮	⋮
Zl > Al+	2	0	1
Distance to NAE	a2 = 2 + 2 +…+2	b2 = 2 + 0 +…+ 0	c2 = 1 + 1 +…+1
Avoidance of RC			
A1 > Z1	0	0	0
A2 > Z2	0	2	0
⋮	⋮	⋮	⋮
Am > Zm	0	2	0
Distance to ARC	a3 = 0 + 0 +…+ 0	b3= 0 + 2 +…+ 2	c3 = 0 + 0 + …+ 0
Total distance to (MA, NAE, ARC)	a1 + a2 + a3	b1 + b2 + b3	c1 + c2 + c3

This account clearly satisfies Inclusion, Closeness Dominance, and Equal Closeness. It also provides a measure of overall closeness to all conditions.

It is also sensitive to the number of violations: if the violations are uniform across cases, all a tie or all a swap, then more violations take us further away from a condition. This is easier to see if we introduce the notion of a *violation vector* that represents how close a theory comes to a condition in different cases. The first value in the vector shows the distance in the first case, the second, the distance in the second case, and so on. If the violation vector for theory T with respect to condition C is (0,

0, x) and for T' it is (0, x, x), where x > 0, then T is closer to C than T' is. But if the violations are not uniform, then one theory can be closer to a condition than another even if the first has more violations. For example, if the violation vector for one theory is (0, 1, 1, 1) and for the other it is (0, 0, 2, 2), then the first theory is closer. If you think this is a problem, you can change the aggregation metric and give more weight to smaller deviations, for instance, by using a function that gives more weight to small deviations (a concave transformation of the distance values in the vector).

Let us now see what the account says about the Mere Addition Paradox, if we consider all possible theories, i.e., all possible rankings of A, A+, and Z. For simplicity, let us again use the toy example with one specific instance of the Mere Addition Paradox, where there are only three specific alternatives to consider, A, A+, and Z.

Fig. 8

	A+ is better than A	Z is better than A+	A is better than Z
0	T1: T2: T3: ***A+*** **A+** Z ***A*** Z **A+** Z **A** **A**	T3: T4: T5: **Z** **Z** A **A+** A **Z** A **A+** **A+**	T5: T6: T1: **A** **A** A+ **Z** A+ **A** A+ **Z** **Z**
1	T7: T8: T9: **A,** Z **A,** **A+** **A+,** **A+**, Z **A** Z	T10: T11: T9: **Z,** A A, **A+** **Z,** **A+**, A **A+** **Z**	T12: T13: T9: **A, Z** A+ **A**, A+ **A, Z** A+, **Z**
2	T6: T4: T5: **A** Z **A** **A+** **A** Z Z **A+** **A+**	T2: T6: T1 **A+** A **A+** **Z** **A+** A A **Z** **Z**	T4: T2: T3: **Z** A+ **Z** **A** **Z** A+ A+ **A** **A**

The top-ranked theories in terms of overall distance to all conditions are the theories with only one violation, a swap: T1, T3, and T5 (overall distance = 2), followed by all theories with at least one tie: T7, T8, T9, T10, T11, T12, and T13 (overall distance = 3), and bottom ranked we have theories with two swap-violations: T2, T4, and T6 (overall distance = 4). This result can be generalized to theories that provide *uniform* violations of the conditions: if the theory entails a certain violation in one case (say, A1+ < A1), then it entails the same kind of violation in all cases (Ai+ < Ai, for all i).

So, we have reduced the initial 13 possibilities to 3 –that is always something— but the remaining top-ranked ones are very different (each alternative gets one top-position, one medium, and one bottom). This means that all population axiologies that judge there to be a tie between some of the populations in the Mere Additions are ruled out. In particular, it means that we have ruled out a person-affecting view, according to which adding new people – moving from A to A+ –does not make an evaluative difference. We have also ruled out a view according to which population A is not better than Z, but only equally as good as Z.

Can we break the tie among the remaining three theories? If not, it is unclear how we can be guided to act by these theories. We can't break it by applying the Kemeny-method again, for that will give us the same set of rankings back. Nor can we break it by applying the majority rule, since it leads to a cyclical ordering. (Note that the three rankings comprise a Condorcet-set.)

But closeness is not the only factor that is relevant when we assess a violation. First of all, some violations are *intuitively worse* than others. For example, a violation of Avoidance of RC that says that Z is better than A even if Z has not more total wellbeing than A seems worse than a violation that says that Z is better than A when Z has more total wellbeing because it is much bigger and the wellbeing of its members is almost crossing the ceiling for being just barely worth living. Similarly, a violation of NAE in which the well-off people are dragged down to the level of being barely worth living, like in A+ compared to Z, is worse than a violation in which one population is a Pareto-improvement of another (all people are at least as well off and some are better off). This means that even if two theories have the same proportion of C-violations, one theory can be preferable to the other because its violations are intuitively not as bad as the ones of the other theory. This suggests the following principle

> If T1's C-violations are more severe than T2's, then T1 is in that respect worse than T2 , other things being equal.

Second, some violations are *farfetched* or *unrealistic*, because they involve populations that could exist in worlds that are very far from the actual world. It seems intuitively less worrisome if the violations of the theory involve populations that are very farfetched. This might in part depend on the fact that our intuitions can be said to be less reliable when the target is some very unusual or farfetched scenario that cannot happen in realistic worlds. It might also depend on the fact that it is less problematic if a theory gives the wrong result in farfetched scenarios than in realistic scenarios.[15] As an example of a farfetched violation, consider violations of Avoidance of RC that involve Z-populations that are of such an astronomical size that they are almost not physically possible. So, two theories can have the same proportion of C-violations, but one is preferable to the other because its violations are more farfetched or more unrealistic. This suggests that

> If T1's C-violations are less farfetched than T2's, then T2 is in that respect worse than T1, other things being equal.

With these extra principles at hand we *might* be able to break the tie. Perhaps all theories tied for distance to the conditions have equally unrealistic violations, but one theory stands out as having less severe violations than the others. To have a greater chance of breaking ties, the simple ranking-distance approach must be revised. We could merge closeness with the other factors and go for an 'element-weighted' Kemeny-measure, according to which the alternatives get weighted so that a more realistic violation increases the distance, and a more severe violation increases the distance. Mathematically this can be done in many different ways, but in order to validate the principles we listed about farfetchedness and severity these weights must make the distance function increasing for both farfetchedness and severity. If we move beyond the toy-example and consider cases where the conditions supply rankings of many pairs of alternatives and the theories order all these alternatives, we have a greater chance to find differences between the theories in terms of the kinds of violations they imply. Of course, nothing guarantees that we will find enough relevant differences between the theories; it depends on which set of theories we consider.

We also have a problem of comparing the severity of a violation of a condition across theories. From which perspective should we carry out these comparisons? One option is to be subjective and just take the perspective of the moral agent. However, one might think that how severe a violation is not (wholly) up to each

[15] I am indebted to Gustaf Arrhenius and Hilary Greaves for this point.

agent to decide.[16] Furthermore, in order to compare all theories, we will have to do some trade-off between the *different* kinds of violations; one theory may have less realistic but much more severe violations than another. How should we trade off these features of violations against each other?

Even if these problems can be solved, we may still be stuck with ties where all considered theories have the same overall distance to the conditions. A partial remedy can be to consider other theoretical virtues, such as simplicity and parsimony. Furthermore, we can consider the credences we have in the conditions. Perhaps we have more credence in two of the three conditions, which would speak in favour of the theory that satisfies those two conditions.[17] More specifically, we could weight the distance between a theory and a condition by its credence.

8. Concluding remarks

This is as far as I have come in my thinking (not that far admittedly). I am unsure about how to answer all the questions surrounding how to construct a satisfactory weighted Kemeny-measure. This may provoke a very disconcerting thought: have we embarked on yet another wild goose chase, leading to another impossibility theorem, this time at a higher level?I can't show you that we need not worry about this. But note that there has been a lot of theorizing on weighted Kemeny-measures and there seems to be no known, *very general*, impossibility theorem that the researchers on ranking-distance stumble on.

In my particular application, I need to sort out the trade-off between different features of violations, but perhaps we can give people quite a lot of leeway on how to do this. Except for some general constraints, it is up to the decision-maker to decide on the trade-off between farfetchedness and severity. If the decision-maker is unsure about how to do this in all relevant cases, we can ask her to assign the alternatives some *set* of weights so that we at least get a *partial* trade-off ordering: x is more distant than y if it is more distant on all weight assignments.

There are further issues to be addressed, for recall that the discussion in this paper was premised on some simplifying assumptions. Which questions do we have to face if we lift these assumptions?

Full comparability. If we relax this assumption, we need to be able to compare *gaps* with swaps and ties. Which comes closer to a certain strict ranking? On the one

[16] Thanks to Hilary Greaves for pressing me on this issue.

[17] To determine how much credence we have in a theory we might need to know how the theory explains the value ordering. This means that we need to go beyond the minimalist framework that identifies theories with their orderings.

hand, it seems closer to a tie than a swap, since it agrees with a tie that the ranking is not reversed. On the other hand, it seems to take us further away from both swaps and ties, since a gap denies the comparability of the alternatives in question.[18]

Disjunctive conditions. I have assumed that the conditions provide strict rankings of pairs of alternatives (A+ should be better than A, Z should be better than A+, and A should be better than Z.) But what should we do when the condition provides a *disjunction* of rankings, for example A+ is *either* better than *or* equally as good as A? It seems reasonable to first determine the distance between the theory and each disjunct and then choose the *shortest* of those distances as a measure of how close the theory is to the disjunctive condition. After all, to satisfy a disjunctive condition is to satisfy one of the disjuncts. For the example above this means that a theory that says that A+ is worse than A is at a distance of 1 from satisfying the disjunctive condition, for it only takes one change (from A+ < A to A+ = A) to satisfy one of the disjuncts.

Closeness to transitivity. How do we measure closeness to the transitivity condition? This is actually not a problem for the Kemeny-measure approach. We can ask how many changes it takes to transform a target ranking into a transitive ranking. So, for instance, a violation of this kind, x > y, y > z, x = z, will be closer to the transitivity condition than a violation of this kind x > y, y > z, x < z. The former requires one change (from x = z to x > z) and the latter two (from x < z to x = z and then to x > z).

Closeness to the universal domain condition. This can be measured by the proportion of cases that the theory applies to, at least when we consider realistic cases.

Even if we can avoid another impossibility theorem, we can wonder whether it is worth trying to work out the best weighted Kemeny-measure. We started with the observation that decision-makers are in a hurry and we ended with yet another theoretical puzzle (This is a typical outcome when philosophers try to be practically relevant). Why think this puzzle is easier to solve than deciding which condition to drop?

I think the options are not exclusive. When we try to work out the Kemeny-metric and how to apply it to the paradoxes, we simultaneously consider how worrisome the violations of the conditions are. This evaluation process can make us reassess the plausibility of some condition(s); perhaps they were overshooting:

[18] If, as is claimed by Rabinowicz and Hájek (2022), we can talk about x and y being incomparable but x being almost better than y, then this kind of gap is closer to x being better than y, than to y being better than x.

some violations now seem acceptable. So, engaging in working out the closeness metric can give us ideas about which condition to drop, even if we fail to find a satisfactory closeness metric. Additionally, this work can also give us reason to change our credence distribution in the conditions, which can help us if we want to go for the hedging option. So, a failure to find a closeness metric can have instrumental value for the other approaches to addressing impossibility theorems.

In any case, we are not completely empty-handed as things stand, for we have established the following principles:

- Closeness Dominance
- Closeness Equality
- Inclusion
- Identity
- A greater proportion of violations takes a theory further away than a lesser proportion of violations, other things being equal.
- Swaps take a theory further away than ties, other things being equal.
- Violations that are more severe make a theory worse, other things being equal.
- Less farfetched violations make a theory worse than more farfetched ones, other things being equal.

Together, these principles will give us some limited guidance on how to rank theories. We can already, at least, rule out certain theories. We can already tell decision-makers *not* to use certain theories. This is progress of some sort.

References

Arrhenius, G. (forthcoming). Population ethics: The Challenge of Future Generations. Oxford University Press.

Arrhenius, G. (2021). 'Population paradoxes without transitivity'. In in Oxford Handbook of Population Ethics, Arrhenius, G, Bykvist, K, Campbell, T, and Finneron-Burns, E. eds., Oxford University Press.

Blackorby, C., Bossert W., and Donaldson D., (2005) Population Issues in Social Choice Theory, Welfare Economics, and Ethics: 39. Cambridge University Press

Brennan, G. (2015) 'Liberty, Preference Satisfaction, and the Case against Categories, in Weighing and Reasoning'. In Themes from the Philosophy of John Broome, Hirose I. and Reisner, A. (eds.), Oxford University Press.

Brennan, G. and Baurmann, M. (2006) 'Majoritarian inconsistency, Arrow impossibility and the continuous interpretation: A context-based view'. In C. Engel and L. Daston (eds.), Is There Value In Inconsistency? Baden-Baden, Germany: NOMOS, 93–118

Bykvist, K. (2022) 'Evaluative uncertainty and population ethics', in Oxford Handbook of Population Ethics, Arrhenius, G, Bykvist, K, Campbell, T, and Finneron-Burns, E. eds., Oxford University Press.

Campbell, D. E. and Kelly, J. S. (1994) 'Trade-off Theory'. Papers and Proceedings of the Hundred and Sixth Annual Meeting of the American Economic Association, The American Economic Review 84:2, pp. 422–426.

Cook, W. D. and Seiford, L. M. (1978) 'Priority Ranking and Consensus Formation'. Management Science 24:16, pp. 1721–1732.

Duddy, C. and Piggins A. (2012) 'The proximity condition'. Social Choice and Welfare 39: 2/3, pp. 353–369.

Greaves, H. and T. Ord (2017). 'Moral uncertainty about population axiology'. Journal of Ethics and Social Philosophy 12:2, pp. 135–167.

Hamming, R. W. (1950). "Error detecting and error correcting codes". The Bell System Technical Journal. 29 (2): 147–160.

Kemeny, J. G. (1959). 'Mathematics without Numbers' Daedalus 88:4, pp. 577–591.

MacAskill, W., Bykvist K., and Ord T. (2020). Moral Uncertainty. Oxford University Press.

Rabinowicz, W. and Hájek, A. H (2022). 'Degrees of commensurability and the repugnant conclusion', Noûs 56:4, pp. 897–919.

Riedener, S. (2021) Uncertain Values: An Axiomatic Approach to Axiological Uncertainty, Philosophical Studies 177, 483–504..

Stephen M. Gardiner[1] & Catriona McKinnon[2]

Generationally Parochial Geoengineering: Early Warning Signs of a Basic Threat

'Geoengineering' has come to refer to massive technological interventions into fundamental earth systems on a planetary scale, often with the aim of counteracting human-induced climate change. Despite a burgeoning literature, some ethical issues surrounding geoengineering remain under-analyzed, barely identified, or in effect ignored. We are interested in one such issue, the threat of generationally parochial geoengineering (GPG): geoengineering that is dominated by the narrow, generation-relative concerns of a given generation engaging in the intervention, without due consideration for wider concerns, including especially the interests of later generations. In this paper, we develop the basic idea and identify some early warning signs in the current discourse, focusing on stratospheric sulfate injection, a form of solar radiation management. Our emphasis is on motivating the claim that generationally parochial geoengineering is a threat that should taken seriously at all levels of work on geoengineering, including research, development, and deployment.

[1] Professor of Philosophy at the University of Washington, Seattle. E-mail: smgard@uw.edu
[2] Professor of Political Theory at the University of Exeter. E-mail: c.mckinnon@exeter.ac.uk

> "There is a kind of parochialness in time. How many writers have there been who have expressed the aspirations of their own generation only?"[3]

1. Introduction

Despite a burgeoning literature, some ethical issues surrounding geoengineering remain under-analysed, barely identified, or in effect ignored. In this paper, we explore the threat of generationally parochial geoengineering ('GPG'): geoengineering that is dominated by the narrow, generation-relative concerns of a given generation engaging in the intervention, without due consideration for wider concerns, and especially the interests of later generations.[4] In our view, thwarting GPG ought to be a central concern of both the ethics of geoengineering, and any serious scientific, political or policy discussion. Unfortunately, this is not yet the case: existing proposals for geoengineering research and governance are largely silent on the threat, and some may even encourage it.

This neglect is lamentable, but also sadly predictable. It underscores the difficult context in which interest in geoengineering is emerging. The history of international climate policy is largely one of severe moral failure, which has now led to a climate emergency. Advocates for pursuing geoengineering aim to moderate the crisis, but tend not to dwell on the underlying nature of the problem that leads to it, nor its implications for their proposed solutions. This is a dangerous omission. Specifically, in our view the climate problem is best understood as a severe ethical challenge that Stephen Gardiner calls 'a perfect moral storm' (Gardiner 2011). On this analysis, two of the main drivers of moral failure are serious discrimination against the future (roughly, Gardiner's "intergenerational buck-passing") and the distortion of the ways we think and talk about climate change, often under the influence of narrow, short-term, and self-serving motives (roughly, Gardiner's "moral corruption").[5]

---

[3] 1906 Academy 20 Oct. 391/2, cited by OED.

[4] This is not the first time the threat of GPG has been raised. Gardiner mentions it briefly in several places (e.g., Gardiner 2011a, 2011b, 2017). We aim to flesh out the concept and show that the threat is live in practice.

[5] Gardiner motivates the general idea of moral corruption by drawing on a passage in Kant (306). Based on that passage, he says that "moral corruption is: (a) a tendency to rationalize, which (b) casts doubt on the validity and/or strictness of moral claims, by (c) seeking to pervert their status and substance, and in doing so (d) aims to make those claims better suited to our wishes and inclinations, and (e) destroys the characteristics in virtue of which we respect them (e.g., what Kant calls their "dignity")" (307). He remarks that moral corruption "strikes at our ability even to understand what is going wrong in moral terms, by subverting moral discourse to other (usually selfish) ends" (305). While Gardiner does not take himself to be offering a precise definition of 'moral corruption', he takes his discussion to be "sufficient for present purposes" (Gardiner 2011, 303-307).

In this paper, we pursue the idea that geoengineering policies are at least as vulnerable to these drivers as more conventional strategies. In the first half, we explore the concept of generationally parochial geoengineering; in the second half, we identify some early warning signs in the current discourse in science and policy. Our focus is on motivating the claim that the risk of GPG should be taken seriously at all levels of work on geoengineering, including research, deployment, governance, and institution-building.

2. Context

Roughly-speaking, we take the term 'geoengineering' to refer to deliberate technological interventions into fundamental earth systems on a massive, typically planetary, scale.[6] Currently, such interventions are being proposed with the aim of counteracting human-induced climate change. While numerous techniques have been suggested, we will focus on stratospheric sulphate injection ('SSI'), the proposal to spray sulphate particles into the stratosphere in order to deflect a fraction of incoming sunlight back into space and so moderate anthropogenic warming.

We choose SSI for three reasons.[7] First, SSI is at the centre of current controversies. It is the focal strategy for geoengineering scientists, already prominent in public discussion, and likely to become more so as the climate situation deteriorates.[8] Second, everyone agrees that SSI is a paradigm case of geoengineering. Third, since SSI is a paradigm case, it is plausible that many lessons from our analysis will carry over to other forms of geoengineering (albeit with suitable modifications for differences in salient features and context). For the rest of the paper, then, assume that when we speak of geoengineering, we have SSI in mind.

One reason geoengineering is being discussed is the emissions crisis. The central goal of international climate policy, agreed over thirty years ago at the Rio Earth Summit, is to protect current and future generations against "dangerous anthropogenic interference with the climate system" (United Nations 1992). In Paris in 2015,

[6] Our definition is broadly similar to the Royal Society's (Shepherd et al, 2009), but does not include combating climate change as part of the definition.

[7] The recent literature tends to use the broader term 'stratospheric aerosol injection' ('SAI'). One reason is that some scientists are now actively discussing using particles other than sulphates, including artificial particles especially engineered for the purpose. We prefer to stick with the narrower term, SSI, in part so as to leave open (for now) the question of whether a slide towards the broader term, SAI, raises further issues. Consider two examples. First, we are concerned that some forms of SAI involve introducing novel, artificial particles into a delicate part of the climate system. Second, we wonder about the implications of breaking the so-called "natural analogy" with volcanic eruptions. Among other things, neglecting such differences between SSI and SAI may facilitate moral corruption.

[8] See, for example: Crutzen 2006; Cicerone 2006; Gardiner 2007, 2011; Hamilton 2013; Keith 2014; Preston 2012, 2016; Gardiner, McKinnon and Fragniere 2021; Stephens et al. 2021; Biermann et al. 2022.

the international community declared that this requires "holding the increase in the global average temperature to well below 2°C above pre-industrial levels and pursuing efforts to limit the temperature increase to 1.5°C" (United Nations 2015; 2009; IPCC 2018). Unfortunately, the world is not close to meeting these goals through cutting greenhouse gas emissions, and time is fast running out. Indeed, some authorities claim that there is now 'no credible pathway to 1.5C' (UNEP 2022).

The emissions crisis helps to explain the appeal of radical geoengineering interventions. Advocates claim that SSI in particular can reduce the urgency of the crisis by moderating climate impacts and altering the political landscape. In a raw form, the initial standard arguments include: SSI will cool the planet quickly and relatively cheaply; SSI will "buy time" for more traditional mitigation efforts; in developing SSI we will be "arming the future" by equipping younger generations with technologies that can limit warming late in the century; and SSI provides a "last resort" to deploy in the case of a climate emergency.

Importantly, the initial arguments have been subjected to sustained scrutiny over a prolonged period.[9] Notably, some of the critical feedback aims to block the pursuit of SSI altogether, while some seeks instead to reshape it, particularly in a more just or ethical direction. Either way, we believe there should be a presumption against simply accepting the initial, standard arguments for SSI at face value, without further, more sophisticated development, and in particular without learning from the implications of engaging with the critics.

Despite this, in practice the standard arguments appear to be becoming highly influential as the emissions crisis continues to intensify. Notably, calls for pursuing SSI have now become mainstream in science and policy circles. For instance, recently a major report from the National Academies of Science, Engineering and Medicine advocated for a US research program, and the effort has garnered wider support from editorials in influential venues, such as *Nature* and the *Economist* (e.g., Economist 2021; National Academies of Sciences 2021; Nature 2021) as well as major think tanks, such as the Council on Foreign Relations (Patrick 2022)).

Indeed, we are concerned that SSI may soon become normalized, without much public discussion or serious deliberation, and with only marginal attention being paid to the social and political issues it raises (e.g., Gardiner 2010, 2020; Stephens et al. 2021; Biermann et al. 2022). One sign that this may occur is that large-scale deployment of another class of geoengineering technologies, carbon dioxide removal

[9] Critical voices include: Jamieson 1996; Gardiner 2007, 2010, 2011ab, 2013ab; Hulme 2012; Preston 2012, 2013, 2016; Hamilton 2013; Fragniere and Gardiner 2016; McKinnon 2019, 2020; McLaren and Corry 2021a; Gardiner, McKinnon, Fragniere 2021. The initial arguments for pursuing SSI also have their defenders (e.g., Svoboda 2012; Morrow and Svoboda 2016; Moellendorf 2014; Morrow 2020; Callies 2022).

(CDR), is already being *assumed* in mainstream scientific projections (e.g., IPCC 2018). Notably, this prominence is being given to CDR even though the main techniques being discussed are poorly understood and largely untested. Indeed, most are *highly speculative*: they either do not yet exist, or are in very early stages of investigation. Given this, large uncertainties hang over whether a massive deployment of CDR is likely to be feasible, on what timescale, and with what risks (e.g., Burns and Nicholson 2017).

3. Generationally Parochial Geoengineering

Over the last three decades, an established academic literature on the ethics of SSI has emerged.[10] This literature identifies a wide range of concerns. Prominent issues include that unethical forms will emerge that encourage or embody serious injustices, including procedural injustice, substantive injustice, and injustices centered on a lack of recognition of diverse values and populations. Some more specific ideas are that actual SSI is likely to be politically illegitimate, encourage moral hazards (such as mitigation deterrence), increase militarization, actively facilitate (or enhance the potential for) oppression by powerful actors, and pose risks to future generations.[11] In our view, this literature is valuable. However, some parts remain underdeveloped and often underappreciated. In this paper, we highlight the threat of generationally parochial geoengineering, a specific kind of intergenerational injustice that involves the *way* in which geoengineering is pursued, including the *kinds* of research programs, interventions or policies that find favor. In highlighting the threat of GPG, we hope to establish it more firmly as a central concern for geoengineering ethics and policy. While we do not believe that recognizing the threat should diminish or supplant other concerns, we also maintain that GPG should not be marginalized or set aside in favor of them. SSI raises many challenges. GPG should be considered in the core group. One reason is that minimizing discussion of GPG encourages intergenerational moral corruption.

[10] For a few examples, see previous footnote. Pamplany et al. 2020 provides a useful (though incomplete) survey of the literature, covering more than three hundred sources from 1996-2020; Fleming 2010 puts the discussion in a wider context. Notably, sometimes the discussion is subsumed under broader terms such as 'solar radiation management' (SRM), 'solar geoengineering' (SE), 'climate engineering', or 'geoengineering' in general. Even then, SSI is typically the focal point of discussion.

[11] Pamplany et al. identify two waves of research so far. The first is characterized by broader, multi-dimensional analysis; the second tends to focus on particular criticisms, most of which were already raised in the first wave, albeit often in less detail. Thus, the concerns listed in the text characterize both waves.

3.1. 'Parochialism'

In general terms, an approach or attitude is parochial if it is "limited or provincial in outlook or scope", or (more robustly) "concerned with only narrow local matters without regard for more general or wider matters".[12] In common parlance, 'parochialism' has come to have negative connotations, signaling a kind of disapproval. For example, the *Collins English Dictionary* characterizes the primary meaning of 'parochial' by saying: "If you describe someone as *parochial,* you are critical of them because you think they are too concerned with their own affairs and should be thinking about more important things". This negative sense of 'parochial' is the one we have in mind in putting forward our term, 'generationally parochial geoengineering' (GPG). GPG refers to geoengineering that is dominated by the narrow, generation-relative concerns of a given generation engaging in the intervention, *without due consideration* for wider concerns, including especially the interests of later generations of human and nonhuman life over the longer term. Thus, GPG involves *a lack of due consideration or proper regard* for wider matters, and in particular substantive and procedural ethical concerns. For instance, under many circumstances, GPG will entail a neglect of rights, justice, and well-being.

We see GPG as one member of a family of troubling parochialisms that could take hold in the geoengineering context, including nationally-parochial geoengineering, corporately-parochial geoengineering, culturally-parochial geoengineering and anthropocentrically-parochial geoengineering.[13] While all these various parochialisms deserve more attention, here we focus on generationally parochial geoengineering. One reason is that GPG remains underexplored, especially in the context of actual climate policy. Another reason is that many of those who stand to be damaged by GPG are politically disadvantaged and indeed often invisible, since they are not

[12] The generic form of parochialism can sometimes be ethically defensible, for instance when (i) the agent has a special responsibility towards the narrow concerns and (ii) others can be relied upon to promote and protect the wider ones. Thus, it is sometimes reasonable for local politicians to prioritise what will most benefit their constituents over (say) national goals. Still, much depends on assumption (ii). Without (ii), it is at best unclear whether those with special responsibilities are absolved of wider claims that might be made on them.

[13] Most obviously, *Nationally Parochial Geoengineering* is geoengineering that is concerned only with the narrow, national concerns of the country then engaging in geoengineering. For instance, country X may deploy the form of SSI that it believes best protects its own agricultural systems, but be indifferent to the severe droughts this inflicts elsewhere. Similarly, *Culturally Parochial Geoengineering* is geoengineering that is concerned only with the narrow culturally specific priorities of a particular dominant culture. For instance, forms of SSI may emerge that are oriented toward preserving Western consumer society, but indifferent to the needs of traditional tribal peoples and values that are central to their ways of life. One example might be SSI aimed at protecting only against a high temperature threshold (e.g., 2-2.5°C), where that does nothing to prevent the loss of small island nations or the Amazon rainforest.

yet born, or are too young to resist. Thus, we in the current generation have a special responsibility to raise concerns on their behalf that they cannot.[14]

3.2. An Illustrative Example

To illustrate GPG, let us sketch a stylized example that we take to be a paradigm case.

> *Elders-First Geoengineering*: Suppose that the current generation of political, economic, and scientific leaders in a powerful, advanced, consumerist society primarily represents and is constituted by those over fifty-five. Call this leadership class "the Elders", and the country they lead "Boomerland".
>
> The Elders become concerned about climate change and the deteriorating situation. Given this, they decide to push ahead with SSI research, aiming to develop a deployable technology as soon as possible. However, rather than trying to limit future warming as such, the core intentions of the Elders are (first) to protect *their generation* from impacts that arise during their own lifetimes, and (second) to continue to enjoy their high consumption lifestyles. In other words, the only negative climate impacts that concern the Elders are those occurring over (say) roughly the next few decades. Consequently, they are drawn to the idea of a relatively short-term "technological fix" that does a reasonable job of holding off the worst effects of climate change for around (say) forty years. They therefore support restricted research and development targeted at interventions that fit this profile, and some forms of SSI look promising.
>
> Unfortunately, it turns out that these forms of SSI also pose extreme risks to younger generations, including some alive late in this century but especially those around in the 22nd, 23rd and 24th centuries. These risks are more severe even than those of substantial climate change (e.g., because SSI encourages runaway climate change, or due to the possibility of termination shock amidst climate breakdown). However, the Elders are either indifferent to these later risks, or at least not highly motivated to prevent them. Notably, this is so even though other, more ethically attractive policies are available, including some that involve SSI done in a different way, or as part of a different portfolio of policy options. Alas, the Elders are simply not interested in alternatives, or in ethics; e.g., they are generationally ruthless, or self-absorbed, or cowardly (see below our discussion of the possible roots of GPG).

[14] The same holds for anthropocentric parochialism and some kinds of cultural parochialism.

> The Elders use their disproportionate power in society to push forward development of their short-term "geoengineering fix", and Boomerland ultimately ends up deploying their preferred form of SSI. As it turns out, this deployment does hold off the worst impacts until late in the century (as intended); however, it then unleashes increasingly catastrophic impacts over the next three centuries (as foreseen). Thus, while it protects the Elder generation, the policy ultimately causes suffering on a global scale, and in ways that are clearly ethically indefensible (by assumption). Humanity ultimately survives, but only barely.

The purpose of the stylized *Elders-First* example is to sketch a paradigm (and thereby uncontroversial) case of a geoengineering scenario that would be rejected by any reasonable moral or political philosophy. For instance, one might say that, other things being equal, any generation acting as the Elders do is (at best) reckless with respect to the basic rights, interests, and needs of future people, and (at worst) engaged in deliberate aggression against them. We expand on this thought shortly. Before doing so, let us explore the richness of the notion of GPG.

3.3. A Big Tent

'GPG' refers to geoengineering that is dominated by the narrow, generation-relative concerns of a given generation engaging in the intervention, without due consideration for wider concerns, including especially the interests of later generations of human and nonhuman life over the longer term. This characterization leaves much open. Since instances of intergenerational parochialism might vary along various dimensions, GPG is a "big tent". To illustrate this, consider just four such dimensions.

First, GPG might involve different *agents*. Some salient possibilities include that the parochial generation may be: the current generation of political leaders (e.g., those controlling the corridors of power over a few election cycles, or those presiding over a despotic reign); or the current generation of social decision-makers (e.g., those in positions of power aged 40–80, who largely determine the political leaders); or the current generation as a whole (e.g., in a highly democratic state).

Second, the *time-frame* over which generational parochialism is operative can vary. For instance, those over 55 may favor SSI that holds off the worst impacts of climate change for, say, 40 years, whereas those over 35 may prefer 60 years, and those over 15 years old may prefer 80 years. Alternatively, perhaps the relevant group is all those alive now, who would like to protect their own children and so prefer technologies that defer truly nasty climate damages for 100–150 years. Although

these time-frames are very different, the basic dynamic of generational parochialism remains.

Third, the *epistemic* conditions of the relevant agents might diverge. For instance, while generationally parochial SSI may arise under *active awareness* of consequences for the longer-term future, it may also occur under limited knowledge or even ignorance. Notably, these do not automatically excuse the parochialism. For example, the epistemic deficits may themselves result from indifference, including a failure even to investigate the relevant implications of SSI beyond a few decades. Under many circumstances, this would amount to intergenerational negligence by the current generation.

Fourth, GPG may have various *roots*. For instance, one salient possibility is that a given generation is *ruthless*: it strongly prefers to advance its own interests, and does not care about the burdens this imposes on future generations. Alternatively, a generation may be *self-absorbed*: it is so focused on what happens to itself, in its own time, that it fails even to seriously consider what happens to its successors. Another potential root is *cowardice*: the generation lacks the moral courage to make the necessary choices on behalf of the future, perhaps because these choices would demand more of it than geoengineering. Importantly, all of these possibilities are dangerous, not least because they open the door to injustice, perhaps of severe forms.

Considering the various dimensions of temporal parochialism, GPG clearly covers a wide range of possible scenarios. Putting a few variables together, we can identify some variants of the paradigm case that may serve as useful touchstones. Among the more obvious are:

- *Last Dance Politicians*: A senior generation of political leaders (e.g., those aged 65 and over) pursues SSI that it hopes will hold off the worst impacts of climate change "on their watch" and immediately afterwards (e.g., for 20–30 years), even with the awareness that their intervention will likely make the situation much worse thereafter.

- *Greedy CEOs*: The current generation of corporate leaders pursue SSI that would hold off the worst impacts of climate change for a few decades and so preserve their power and profits, showing no interest in the impacts thereafter.

- *Generational Elite Capture*: The current generation of social decision-makers (e.g., those in positions of power aged over 45) pursues SSI that holds off the worst impacts for 50 years and fails to investigate the implications for later generations.

Among the less obvious, but still highly salient, scenarios are:

- *Those We Love Now*: The current generation of social decision-makers (e.g., those in positions of power aged over 45) pursues SSI that holds off the worst impacts for itself and its children (say, for 80–120 years), but fails to investigate the implications for later generations.

- *An Unholy Alliance Against the (Further) Future*: Young adults and particularly vulnerable communities cooperate with older generations in backing a climate policy (a) that pursues SSI to hold off the worst in the short-term when the older generations are around to benefit (e.g., 30–40 years), (b) in exchange for medium-term adaptation measures that help to protect the current young over much of their lifetimes (e.g., 50-60 years), (c) while accepting that this approach worsens impacts in the further future (e.g., after 80 years).

- *An Intergenerational Arms Race*: A succession of generational agents each seeks to postpone negative climate effects that would fall on themselves by shifting the worst impacts to the future. The cumulative effect of all this buck-passing is to compound those negative effects on some generations in the further future, dramatically driving up the risk of eventual catastrophe, perhaps to the point where it is inevitable (cf. Gardiner 2011b, chapter 6).

3.5. Normative Roles

A further aspect of the richness of the GPG analysis concerns the variety of roles it can play in normative contexts. To begin with, in our view, generational parochialism is a *generic ethical challenge*, to which any ethical tradition will want to respond. Notably, neither the definition nor the cases mentioned above specify the precise normative content of the ethical concern raised by GPG (e.g., by fleshing out the central idea of failing to give 'due consideration' to wider, intergenerational concerns). For instance, the worry that GPG identifies is not indexed to a specific set of normative concepts (e.g., human dignity, impartiality, rights, equality, sufficiency, just savings), nor associated with a particular ethical tradition (e.g., Kantian ethics, utilitarianism, virtue ethics, ethics of care). Instead, the background idea is that GPG is and ought to be a cause for concern *whichever* normative framework or ethical tradition one favors.[15] For example, GPG can be objected to as a failure of impartiality, a lack of respect for future human rights, or a failure to secure conditions

[15] Some may reject this burden. Perhaps they reject intergenerational justice or ethics in general (e.g., because they maintain that self-interest is the only or overriding concern), or perhaps they believe that

sufficient for future people to flourish, or all of these at once. Deciding which is the right approach is an important task within intergenerational ethics, but the concept of GPG does not presuppose a particular answer. This is a matter for deeper theory.

We cast our net widely for a reason. We take the ability of an ethical framework to confront the tyranny of the contemporary to be a *condition of adequacy* for that framework, and one through which rival traditions might be compared and judged. In other words, we see avoiding generational parochialism (here in the context of geoengineering) as a test for ethical theories in the same way that endorsing universal suffrage is a test for theories of democracy. Approaches to ethics will want to show that their central concepts are well-suited to make sense of and neutralize the intergenerational threat. If a given approach directly encourages GPG, then it is in trouble.[16]

Nevertheless, our concern in this paper is elsewhere. We aim to make the case that there are good reasons to treat GPG as a *live threat*: given how actual discussions of SSI have been evolving, there are serious concerns that the situation is primed for GPG to emerge. Given this, our focus is on what one might call "providing *guidance against temptation"*. We wish to alert relevant parties (e.g., scientists, policymakers, the public at large) to the general threat of GPG, to help them recognize ways in which more concrete practices may encourage GPG, and to point out some places where generational parochialism may already be creeping in. Our hope is that increased awareness will act as a *first line of defence* against GPG, and so help to forestall or pre-empt the worst excesses. That being said, we are not optimistic that awareness will be enough by itself. In our view, further, institutional defences will ultimately be required to check GPG, some of which are likely to involve radical shifts away from the status quo (e.g., Gardiner 2014a, 2019; Gonzalez-Ricoy and Gosseries, 2016; McKinnon 2017, 2021; see below).

This focus on providing guidance against temptation has implications for how one should understand the kinds of evidence we are looking for, and how that evidence should be treated, when we turn to early warning signs. Since we intend simply to make the case that there are good reasons to regard GPG as a live threat, we are looking for grounds for *reasonable suspicion* that there is a risk of GPG emerging. Identifying early warning signs can serve to put us on our guard by suggesting that there is *an initial case to answer*.

Importantly, this focus is very different from that of *convicting* particular agents

---

it is never inappropriate for the current generation to ignore the concerns of other generations. We find such views unpalatable. In any case, we leave them aside here.

[16] This claim is not inconsequential. For instance, some will argue that conventional cost-benefit analysis – i.e., that based on projections of current market prices, and employing standard positive discount rates, such as 5% - fails the test.

of GPG. Although such criticism might be of interest in the future, especially to future generations, this is not our purpose here. Rather than censure, our primary aims are redirection, resistance and reform. It is largely our generation that needs to act, and not succumb to the temptation of intergenerational buck-passing. Since we are the ones implicated, much of the point of identifying GPG is to influence our behaviour, by showing us what is at stake and putting us on our guard, and to prompt institutional reform. Conviction and censure are not necessary to achieve this; moreover, they may end up being beside the point, as they are likely to come too late.

One important consequence of the focus on redirection and reform is that we should not be fixated on satisfying the high standards of proof that are often taken to be needed for the purposes of conviction or blame. Instead, we can be content with much lower levels of evidence. In particular, we are interested in evidence that is *sufficient to activate a duty to protect* the future against the predictable threat of GPG. In light of this, we should not, for example, direct our attention to trying to establish that GPG is present "beyond reasonable doubt". Instead, noticing that there is a reasonable suspicion of GPG, or even merely a lurking danger, can be enough. This level of awareness should be sufficient to put those working on geoengineering interventions on their guard to look out for signs of GPG, and help them to develop a sensitivity for problematic practices or assumptions that may facilitate GPG. For example, it can make them alert to potential blind spots or implicit biases in geoengineering research or policy. Reasonable suspicion of GPG also helps to ground the wider case for institutional reform, and to suggest that a specific goal of such reform should be to confront GPG. In this way, there are parallels between raising awareness of the threat of GPG and confronting other social problems, such as subtle forms of gender bias or institutional racism.

Our fourth point is that (perhaps surprisingly) investigation of GPG can be worthwhile even if our initial arguments fail to establish that there is already a case to answer. For active engagement with the possibility of GPG itself heightens awareness in ways that make it less likely that generational parochialism will emerge in the future. Once relevant groups, such as scientists, policy professionals, government officials, institutional reformers, and the general public, start checking geoengineering proposals for the possibility of generational parochialism, this encourages a positive feedback loop whereby GPG is more likely to be confronted at all levels, including at the earliest stages and in foundational assumptions. Thus, paradoxically, it can turn out that highlighting the threat of generational parochialism can be successful in providing guidance even if no actual instances of GPG are ever shown to have occurred. Indeed, in some ways that is the best-case scenario for the project as well as for humanity.

Our fifth and final point about the normative character of GPG and our interest

in it is more substantive. Despite all this modesty, it is true that our own concern about GPG emerges from a particular mindset about the appropriate ends and framing of geoengineering policy. Most centrally, in the background is our belief that ultimately (if pursued) SSI should be seen and assessed as a global, intergenerational, ecological, and ultimately ethical project: one aimed at protecting the interests of humanity (and nonhumans) at large across generations, in accordance with appropriate ethical norms, including norms of justice.[17] This mindset invites the further claims that research and governance should be developed in an ethically responsible way, keeping the global project in mind. Among other things, this suggests that research should also be aimed at protecting the concerns of humanity at large across generations, and that governance should be appropriately responsive to the interests and rights of people globally and intergenerationally (e.g., Gardiner and Fragniere 2018). In light of all this, a focal question for us is "How would future generations view the current pursuit of SSI?"

In the following sections, we turn to this question. Before doing so, however, we want to note one caveat. Our endorsement of a global, intergenerational, and ecological perspective on the appropriate aims of geoengineering is very different from an alternative mindset that claims that geoengineering should be seen as a nationalist project. We do not begin from the place of "America First SSI" or "China First SSI" or "Russia First SSI", and so on. Presumably, it is *possible* that such alternative beginnings may lead to ethically-acceptable outcomes at the global and intergenerational level. However, we are not focused on mere possibility, but on *plausibility*. Importantly, such plausibility would need to be shown, not simply assumed.

Sadly, we are pessimistic. One reason arises due to concerns about competing geoengineering interventions and counter-interventions, multiple-invocations of rights of self-defence, and the prospect of a geoengineering arms race that might be even more dangerous than severe climate change itself (e.g., Gardiner 2013a). However, another reason (particularly salient for this paper) is that nationalistic geoengineering is also vulnerable to GPG. It is not hard to imagine that some nationalistic forms of SSI may be beneficial for the first generation or two of (say) Chinese, Russian, or American leaders deploying it, but much worse for future generations *of those same peoples*. Thus, true nationalists – those who genuinely care about the *long-term* interests of their own countries, and not just about a few fellow nationals alive now – should also have serious concerns about GPG.

[17] See, for example, the first Tollgate principle for governing geoengineering (Gardiner and Fragniere 2018).

4. Specific Early Warning Signs

Let us now look at some specific ways in which the existing scientific and policy literature may encourage GPG, including by obscuring, disguising, or even actively facilitating it. Again, our focus will be on suggesting reasonable suspicions and a case to answer (not on conviction or blame). Again, the overall aim is to encourage a mindset fit to minimise the risk of GPG at various levels (including research decisions, norms, and institutional reform).

4.1. Research

We begin with the framing of research questions in various venues, including models, scenarios, and so on. For instance, climate and earth system models simulate interactions between the various drivers of climate change in order to heighten understanding of the climate system and project future climate change; geoengineering models do the same thing for SSI. Similarly, integrated assessment models simulate interactions between physical and social systems. Almost all current research on SSI takes the form of modelling and scenario building. Yet several common features of these exercises give cause for concern.

Sign 1: Short Time-horizons

The first is the number of years into the future for which models are typically run. The effects of SSI are likely to play out over a very long time-period, of the order of at least many decades, probably several centuries, and possibly thousands of years (e.g., IPCC 2014, 73–74). Yet the overwhelming majority of models have a much shorter time-horizon. For instance, in physical science the models typically focus only on the *next 10-50 years* when estimating the impacts of SSI (Kravitz and et al 2014; Eastham et al. 2018) and a few extend the horizon *only to 100 years* (e.g. Moreno-Cruz and Keith 2013).

This emphasis on short time-horizons also appears to be present in common choices of scenarios for policy analysis. For example, the editors of a recent collection on geoengineering scenarios – two leaders in the field – assert: "to be useful, ... creators and users must judge the scenario, or a similar pattern of events, as sufficiently likely to merit their attention and consideration in planning", yet they go on to say that **"***all [scenarios* in this collection] were set in the year *2040*" (Parson and Reynolds 2021, 5–7, emphases added).[18] Another example comes from conventional economic analysis. Referring to the dominant economic model (DICE), a

[18] They add: **"**This date was chosen to be near enough that scenarios are not dominated by vast technological or socio-political transformations and their relevance for near-term decisions is clear, while also being distant enough that greatly strengthened social and political forces promoted solar

recent paper on the economics of geoengineering states: "As with most applications of DICE, we are not interested in the very long run" (Belaia et al. 2021).

These choices of relatively short time-frames for research in science, economics and policy provide *prima facie* evidence that mainstream efforts to understand SSI are likely to be preoccupied with impacts on current adults and, at most, their children. It is easy to see why a focus on short time-frames encourages GPG; one might even say that it is a hallmark of GPG. By contrast, an ethical approach to geoengineering appears to require projections over much longer time-frames. If SSI is to have major impacts over many generations across the entire planet, how can research time-frames of only a few decades or even a century be appropriate? Surely there is a case to answer.[19] (See section 4 for further discussion.)

Importantly, there is also reason to think that a longer-term perspective might make a substantial difference. Notably, a recent publication that operated over the much longer time frame of a thousand years suggested a major effect on La Nina events which the lead author, Dr Abdul Malik, said would "strongly impact temperature, precipitation, floods and drought patterns across the globe" (Malik et al. 2020). As a result, Professor Joanna Haigh, co-author and former Co-Director of the Grantham Institute, declared: "The results of this study indicate that solar geoengineering can in no sense be viewed as a sensible rescue plan due to the potential to severely impact on temperature, precipitation, floods and drought patterns across the globe" (Ibbott 2021). We ourselves are not making any such claim – in part because it may be too early to reach such a conclusion.[20] Our point is simply that this new work underscores the importance of considering much longer timeframes than is usual, especially when framing research questions (in modelling, scenario building, and elsewhere). This would be a basic, first line of defence against GPG.

Sign 2: Fast-start Focus

A second feature of mainstream modelling and scenario-building is the choice of starting-points for deployment. In our view, there are good reasons to believe that *any responsible* development of SSI would take at least a few decades of testing, impact assessment, and institution building, and perhaps longer. Consider, for instance, testing. Some claim that "some climate response tests, such as those attempting to detect changes in regional climate impacts, may not be deployable in time periods relevant to realistic geoengineering scenarios". One reason is that "any

geoengineering would be plausible" (Parson & Reynolds 2021, 7). For discussion, see section 4 below.

[19] Some in the geoengineering research community have argued that a short- to medium-term focus in modelling is preferable to a longer-term focus in order to generate information needed by lower- and middle-income countries in their adaptation efforts (e.g., Nissan 2019). Our response is that such modelling can perfectly well co-exist with modelling that has a much longer-term focus.

[20] For instance, this is only one study and involves a large forcing (of quadrupling C02).

deployment scenario in time scales relevant to averting the 1.5C or 2C targets would likely have to proceed with low certainty about regional impacts" (Lenferna et al. 2017).[21] Given the issues surrounding responsible deployment, there are reasonable grounds for saying that, other things equal, *any deployment coming in the next couple of decades is likely to be irresponsible, scientifically and ethically*. Instead of a well-considered intervention backed by the best science, such a deployment is at considerable risk of becoming a *high-stakes, high-risk gamble* in a situation characterized by high levels of ignorance and uncertainty.

More generally, we have some concern that requirements for responsible deployment may turn out to be sufficiently robust that they call into question the very possibility of ethically attractive or even minimally decent forms of SSI becoming available on a reasonable timescale. For instance, although being ready to deploy SSI in a responsible fashion in 2100, 2150 or 2200 would presumably be a major scientific and social achievement, it would not answer the purpose for which many are advocating it now: to avoid breaching the 1.5°C, 2.0°C and other thresholds this century.

A number of reasons underlie such worries. Let us highlight two. The first is that research is still in an early stage of development, such that models remain quite primitive in comparison to the intervention being proposed. For instance, until quite recently, most of the modelling that had been done was essentially of "turning down the sun": reducing incoming radiation at a uniform rate. This is some distance from understanding human attempts to inject aerosols into the stratosphere, and the interactions of those attempts with overall Earth systems over a long period of time. In the last few years, models have moved forward to examine some aspects of injections themselves (e.g., how it matters whether they are done at the equator or other latitudes). Nevertheless, a robust literature is yet to emerge on key issues (e.g., NASEM 2021, chapter 6), such as realistic interventions in the stratosphere at relevant scales, their interactions with broader systems (e.g., other parts of the atmosphere, terrestrial ecosystems), and their long-term consequences. Indeed, this is a central reason for advocating for more scientific research, including research which goes beyond modelling.

The point that there remains considerable work to be done is a simple one, but no less important for that. It underlines the possibility that, ultimately, the gap between where we are scientifically and where we would need to be in order to be justifiably confident in deployment may yet prove so large as to make SSI an unrealistic option over the time period being considered by most policy-makers, and perhaps for even longer.

[21] One of us (Gardiner) is a co-author on that paper.

The second, perhaps more important, reason for concern is that models can take us only so far. At some point, field testing will be needed. Plausibly, this will take at least a couple of decades, and probably significantly longer. Most notably perhaps, establishing a firm evidence base for the safety of SSI is likely to be a challenging task, scientifically and technologically.

One factor is that testing most allow sufficient time for signals to emerge from the overall noise of the climate system. This is especially so if we need to work with a modest injection, rather than a more dramatic forcing, which seems highly plausible given that any actual high magnitude test in the stratosphere will affect people on the ground in significant ways, and so would amount actually to doing geoengineering, rather than simply testing it.

Another factor is that researchers should be interested in the *longer-term effects* of SSI interventions, and robust indicators of these will likely take a while to emerge. Such issues strongly suggest that establishing a firm evidence base will not be a quick process. Yet proceeding to full deployment without a firm evidence base seems very risky, and may even count as reckless and negligent.[22]

Interestingly, concerns about the early stage of research and the demands of responsible testing may be amplified if novel, specially engineered particles will ultimately be used for deployment.[23] One reason is that the move away from sulfates (in SSI) to other, and especially novel particles, may compromise the "natural analogy" with volcanic eruptions, perhaps to breaking point. Another reason is that the implications of introducing novel particles into planetary systems are likely to be more difficult to predict. This is perhaps especially so when one considers the effects on sensitive parts of the atmosphere and on fragile ecosystems on the ground.

Give all this, it is striking that most existing research focuses on quick deployment scenarios. For instance, geoengineering models typically envision a (very) fast

[22] Something depends on how demanding the standards being imposed on testing are, and these may vary in comparison with the risks posed by climate change itself. In our view, this is an ethical issue. However, we cannot pursue it here.

[23] The prospect of creating *specially-engineered, artificial nanoparticles* to inject into the atmosphere was raised by David Keith in 2010: "engineered nanoparticles could exploit photophoretic forces, enabling more control over particle distribution and lifetime than is possible with sulfates, perhaps allowing climate engineering to be accomplished with fewer side effects" (Keith 2010). A 2018 article from Keith's group considers manufacturing "engineered micron-scale particles" with "high radiative efficiency", perhaps "coated with a thin (<10 nm) metal layer" (Parker, Horton, Keith 2018). A 2021 article from another group states: "Even though aerosol injection into stratosphere is one of the most promising solar geoengineering techniques, sulfate aerosols, which are suggested for such an application, show significant drawbacks such as infra-red (IR) absorption and ozone degradation. The development of new materials for such application that would exhibit substantial up-scattering, with non-IR absorption to allow a cooling effect are needed. Here, a novel composite material comprised of diamonds dispersed in a silica aerogel network is investigated and compared to pure silica aerogel." (Vukajlovic et al 2021; emphasis added).

start for SSI, in *only 10-25 years*. One reason for this is that, back in 2010, the original paper from the influential Geoengineering Model Comparison Project ('GeoMIP') assumed deployment would begin in only ten years, in 2020:

> "[The experiment] assumes an RCP4.5 scenario...but with additional stratospheric aerosol added *starting in the year 2020, which is a reasonable estimate of when the delivery systems needed to inject the aerosols might be ready.*" (Kravitz et al. 2011, 164, emphasis added)

The lead author, Ben Kravitz, tells us that this paper had significant influence on modelers and high-level reports:

> *"Numerous climate modeling studies have since begun their simulations in 2020* thanks to GeoMIP's precedent. Many of these geoengineering studies that show a start date of 2020 *are highlighted in reports at national and international scales."* (Kravitz 2020)

As we have indicated, we believe that the timeframe of a mere decade was unrealistic for responsible SSI back in 2010, and there are good reasons to think a fast-start focus remains so today. Kravitz has subsequently been admirably frank about the problems with the decadal modelling, given the state of the science. He also worries more generally about the framing effects, especially in influencing policymakers: "statements from the world's largest geoengineering research effort influence how ideas are shaped and discussed, not just among the scientific community, but also in society and politics". He cringes at the thought that they "might be used as part of a justification for any potential deployment" (Kravitz 2020).

Still, the fast-start focus continues to be present in the literature.[24] For example, a recent article from a top modelling group focused on 2035, choosing it as the start date for most scenarios, and so *only a 13-year time-frame* from publication. For the sake of assessing sensitivity, it also considered 2045 for other scenarios, and so a *23-year time-frame* (MacMartin et al. 2022). Similarly, an assessment of SSI with the goal of protecting the West Antarctic Ice Sheet published in 2015 envisioned deployment in 2035, which was then a *20-year window* (McCusker et al. 2015). In short, it seems common – in fact, the norm – to model for SSI starting in just 10–25 years.

[24] This may be simply because the papers were written before Kravitz's warnings. Moreover, again, we are not aiming to blame researchers, but only to point out how the state of the discussion tends to encourage GPG.

This makes us worry about the lurking threat of GPG. The fast-start focus seems puzzling if one were intent on pursuing an ethically responsible form of geoengineering that aims at the good of humanity as a whole over the very long-term. For one thing, concentrating on a timeframe of only a decade or two to deployment seems highly ambitious given the likely constraints on responsible development coming from (among other things) the need for testing, impact assessment, and institution-building mentioned above. Even more importantly, it is odd to focus *solely* on a 10–25 year window. In general, our recommendation would be that a genuinely intergenerational geoengineering research program should consider a *range* of time-frames for deployment, stretching out into the future. For example, such a program would take seriously preparing for deployment in different time periods, such as 2050-60 or 2070-80 or 2090-2100 or 2110-2120, as well as 2035-2045. Similarly, although investigating SSI to protect the 1.5°C limit makes some sense, it seems problematic, given the state of the science, to make it the *only* scenario considered. After all, perhaps by the time responsible SSI is likely to be ready for deployment, 1.5°C has already been left behind. Thus, an ethical research program would also consider SSI at different temperature thresholds, such as 1.7°C, 2.0°C, 2.2°C.

Disturbingly, the fast-start focus becomes more plausible under GPG. On the one hand, perhaps fast-tracking deployment by ignoring the need to test, assess and build institutions makes some sense if the overwhelming concern is with protecting a smaller group within the current generation. For instance, such a group may be satisfied to proceed if they have decent grounds to assume that any negative impacts would be manageable for a couple of decades or so, even if they may be catastrophic later on, or if the group is not so concerned about their personal longevity and so willing to "roll the dice".[25]

On the other hand, ignoring pathways to responsible geoengineering would also be intelligible if the pursuit of SSI were being endorsed by the parochial generation only for appearances' sake. Touting geoengineering might function as yet another "dangerous illusion" calculated to give the impression that an older generation is doing something about climate change even as it continues to drag its feet about more conventional changes that would clearly make a difference.[26] Deflecting

[25] Another possibility is that fast-tracking deployment may expose the current generation to higher risks of severe negative side-effects than the future. This might encourage the opposite of GPG: the current generation might choose to take on the burden of such risks in order to protect the future, and perhaps even to compensate (in part) for its own bad behavior in not combatting climate change more effectively earlier and in other ways (for this kind of suggestion, see Gardiner 2010, 293). Still, this scenario seems unlikely under current geopolitical realities.

[26] Gardiner calls Kyoto, Copenhagen and Paris "dangerous illusions" of this sort (Gardiner 2004b; 2011; 2022a).

attention from its failures may be another way to "buy time" for such a generation – not for decarbonization or adaptation, but to hold off the disapproval of the younger generations who will be left carrying the can.

More generally, it is easy to see why *only* SSI with a fast-start focus would be of direct interest to an older parochial generation. Since it seeks to protect itself, not the longer term, techniques that would take multiple decades to develop would not be relevant to a buck-passing generation, even if these technologies held the promise of protecting later generations. Consequently, we might see a parochial generation discourage, ignore or veto research on promising forms of geoengineering which would not be available until, say, 2060 or 2075 or thereafter. Instead, they would push for investment in much more messy and speculative interventions that could be deployed in the next couple of decades. Again, the threat of GPG opens our eyes to many risks.

We conclude that the fast-start focus is sufficient to raise suspicions about GPG, and so to put us on our guard and encourage counter-measures. Fortunately, some of these are relatively straight-forward. For instance, at a minimum, we would suggest that a sensible research program into protecting future generations should aim to model and prepare for other salient possibilities than near-term deployment (e.g., 2035 for 1.5°C), including medium-term deployment (e.g., 2050-2060 for 1.7°C), long-term deployment (e.g., 2070-2080 for 2.0°C) and perhaps very long-term deployment (e.g., post 2100 deployment for 2.3°C). Such an expanded mindset would likely increase the prospects of intergenerationally ethical geoengineering.

Sign 3: Neglect of Maintenance and Exit Strategies

A third early warning sign of potential GPG and possible moral corruption concerns long-term maintenance and the need for an exit strategy. Many current geoengineering advocates argue for SSI on the grounds that it will "buy time" for emissions reductions by "shaving the peak" of climate impacts (e.g., Keith and MacMartin 2015). This rationale implicitly assumes that the climate intervention will be maintained for at least several decades, and perhaps centuries, but then ultimately be wound down. Given this, it is striking that little work has yet been done on what these pathways might look like. Instead, while most appear to presume a phaseout, they do not actually model it (e.g., a prominent research group reports that "*only one [paper*] simulates a deliberate gradual phaseout to a warmer world" (MacMartin et al. 2022, 1–2 of 9; emphasis added).) This creates a situation where, in effect if not in intent, the models typically assume that SSI will be ongoing, continuing indefinitely into the future.

Again, this situation would be surprising under ethical geoengineering, but becomes deeply worrying considering the threat of GPG. A current generation

intent on protecting itself and indifferent to the longer-term future would not be motivated to explore how to phase out SSI, if it assumed drawdown would only occur long after it had departed the scene.[27]

As well as the general issue of phaseout, there are some more specific concerns about the focal modelling scenarios and the issues of actively managing SSI, especially for the long-term. A prominent research group tells us: "few papers … have considered a temperature target lower than that at the start date", while "none explore the dependence on the assumed start date" (MacMartin et al. 2022, 1). Both points are concerning, given the risk of GPG. The first does not seem to take seriously enough the idea that global temperature may substantially overshoot mainstream targets, for instance while the testing, impact-assessment and institution-building necessary for responsible SSI is being developed.[28] The second *assumes away* the issue that perhaps the best start date for SSI aimed at the overall intergenerational good of humanity differs from that which would be best for the current generation.

Perhaps the most important concerns, however, are around the potential for serious risks associated with the maintenance of, and ultimate exit from, SSI. The essential role of SSI is to *mask* warming by preventing the accumulation of greenhouse gases from having its full effect. This implies that if SSI is masking a substantial temperature rise, it cannot be safely stopped until the excess greenhouse gases are removed. Thus, substantial SSI must be maintained over a considerable period of time. The reason is simple: if the SSI "mask" is taken away, the planet's temperature will swiftly "bounce back" to the level it would have been absent the intervention. This threat is known as "termination shock" (e.g., Parker and Irvine 2018; Rabitz 2018).

The term 'shock' is employed for a reason. The change would be relatively quick. Current wisdom suggests that the particles injected in the stratosphere (the "mask") would wash out in 6–18 months. Thus, exposure to normal levels of solar radiation would resume within a couple of years, and exposure to the full effects of the rebound within 10–15 years. This kind of rapid warming would likely have *much worse* impacts even than the gradual climate change that the SSI is attempting to block. Moreover, if the masking effect is large, the magnitude of the shock resulting from withdrawing SSI will also be high. For example, if the SSI were holding off a

[27] MacMartin et al. 2022 consider very short deployments of only a few decades, with SSI to be wound up late in this century. This is laudable from the point of view of prompting modelers to think about exit strategies. Nevertheless, it is not clear why it should be the only scenario to be considered, or among the most likely. For discussion, see Gardiner & McKinnon, in preparation.

[28] A reviewer reminds us that researchers are clearly concerned about the risk of overshoot more generally, and often say that it motivates their work. Our observation is more specific: given the risk of overshoot, it is surprising that lower temperature targets are not prominent. This observation encourages worries about GPG.

global rise of 2–3 degrees, then withdrawing it suddenly would see that materialize very quickly by climate standards and human standards.

Most commentators recognize that termination shock is one of the most serious risks associated with SSI, and some believe that it poses such a large threat that we should not seriously consider this kind of geoengineering. Some of the reasons are scientific or technical. People doubt that we can develop or fine-tune SSI sufficiently quickly to a reasonable level, and so fear that the threat to the future of proceeding is too high. Other reasons are political: many are highly skeptical that humanity would develop the kind of governance for SSI that would be resilient enough to provide a decent level of protection against the kinds of failure (whether accidental or intentional) that would result in termination shock (McKinnon 2020). Even if one has faith that *eventually* humanity could achieve this, to assume that we could do so quickly – within the next few decades – is worrying.

Termination shock is explored to some extent in the scientific literature. Still, how to address it, and how to ramp down more gradually remains underexplored. Similarly, "no papers include scenarios that explore the effects of a temporary interruption or other deployment inconsistencies ..." (MacMartin et al. 2022, 1-2). While all of this is worrying, it would be sadly predictable under GPG. Again, it is highly plausible that an older generation focused on protecting itself would not be too concerned about the need for long-term maintenance or an exit strategy. Evidence that SSI is a better bet than climate change over a couple of decades would probably be enough.

A further, more general worry also underlies concern about maintenance and exit. The "buying time" strategy assumes that SSI will be deployed only for a limited period while rapid decarbonization is occurring. However, this may be a bold assumption, and the relevant time-period is uncertain. One issue is, of course, ongoing political inertia around addressing the underlying causes of climate change, and particularly the global economy's heavy reliance on fossil fuels. SSI that is even moderately successful may encourage further procrastination and delay (e.g., the "moral hazard" worry and its cousins)[29]. Moreover, it may do so even as continued intervention becomes more and more risky as it masks ever larger temperature increases.

Another issue is that most proponents of the "buying time" strategy assume that the main way humanity will wean itself from SSI is through directly removing greenhouse gases from the atmosphere, especially through carbon dioxide removal on a massive scale. Yet, as mentioned above, that technology is also highly speculative,

[29] Early references include: Gardiner 2007, 2010, 2011a; Hale 2012; Hamilton 2013. For a more extensive list, see Pamplany et al. 2020, 3093-4.

and may not develop as hoped. Notably, this worry is serious enough to have prompted some prominent climate scientists, such as Ray Pierrehumbert, Professor of Physics at Oxford, to reject SSI completely (e.g., Pierrehumbert 2019).

We conclude that a generation focused on GPG would probably neglect long-term maintenance, exit strategies, the potential failure of CDR, and the threat of termination shock. We therefore suggest that, if we are to forestall GPG, questions surrounding these matters should be much more central to geoengineering research and policy.

4.2. Governance

A second area of concern surrounding generational parochialism and possibly moral corruption involves governance and how it is conceptualized.

Sign 4: Status Quo Bias

Our fourth early warning sign is that current policy analysis often involves what later generations may come to see as a *status quo* bias. For example, the collection cited earlier assumes:

> "[T]he general state of world development and geopolitics is described as *broadly similar to that of today*. Present trends of broad world development and relative decline of dominant powers have continued, but there have been no world wars, regime changes in major powers, or fundamental re-alignments of the international system ... There has still been *no significant progress* at developing relevant international governance capacity ..." (Parson and Reynolds 2021, 7–8, emphasis added).

In short, this is modelling for political and institutional business-as-usual.

Unfortunately, in our view some level of transformation of the global system is probably required to govern geoengineering, and perhaps to deal with climate change itself (Gardiner 2014a; 2019; Maltais and McKinnon 2015; Kashwan et al. 2020; McLaren and Corry 2021b). The status quo, then, while of understandable concern to the current generation, may not be of much use to future generations, and indeed may constitute a core part of the problem they face. Given this, a status quo bias is likely to facilitate and encourage GPG.

Sign 5: Underestimating the Task

The fifth early warning sign is that of underestimating the governance task. Notably, some ways of framing SSI seem highly complacent, even to the point that they

"encourage a kind of *hyper-optimism* about SSI that amounts to utopianism" (Fragniere and Gardiner 2016; Gardiner, 2013b); indeed, some governance proposals appear "Panglossian" (McKinnon 2020). To illustrate this, let us identify three forms of complacency that are common in the literature.

The first is *political complacency*. For example, Catriona McKinnon cautions us that we should not promote governing deployment with policies that *simply assume* a background infrastructure that ensures sustained trust, transparency and cooperation between states with histories of conflict, enmity, and espionage. This makes future people hostages of best-case scenarios coming to pass (McKinnon 2020). Similarly, we should beware of proposals that suggest that nation states and other key actors will easily converge on key ethical aims, such as protecting the global poor (Horton and Keith 2016). After all, progress on similar objectives has not usually been impressive (e.g., the "war on drugs"; the UN sustainable development goals).

The second form of complacency is *institutional complacency*. One aspect of this is procedural. Advocates for SSI tend to focus their efforts on pushing for improving geoengineering science. They do not prioritize, and in general pay much less attention to, the need for effective institutions to govern eventual deployment (including maintenance and exit). This is so even though some of the biggest worries about SSI deployment are its likely lack of political legitimacy (e.g., Gardiner 2011b; Morrow, Kopp, and Oppenheimer 2013; Callies 2019; Gardiner et al. 2021), and that it may be "ungovernable" given the current shape of international politics (e.g., Hulme 2012; Hamilton 2013; Biermann et al. 2022). Tellingly, there seems very little political momentum towards serious institutional change thus far. For instance, we do not see an urgent push to prepare robust new global institutions to govern geoengineering, even as support for scientific research picks up. This may suggest that the current generation of decision-makers are not truly serious about ethical forms of geoengineering, but are instead mostly drawn to unethical forms, including GPG.

Another aspect of institutional complacency is more substantive. Often, the kinds of existing mechanisms and venues recommended for governance appear modest at best. Consider, for example, proposals to refer SSI to the United Nations' Commission on Sustainable Development (Royal Society, 2009) or place it under the UNFCCC. Such approaches seem woefully inadequate given the high stakes and fundamental issues involved in SSI. Other suggestions are somewhat more promising, such as referring geoengineering to an *ad hoc* committee of the United Nations' General Assembly (e.g., NASEM 2021, 190), or to the UN Security Council. Nevertheless, there seems little interest in the idea that fundamental political reform may be required.

By contrast, in our view, existing political institutions and legal systems offer little to no protection to future people against GPG; hence, taking the threat

seriously requires a strong focus on governance, and one that likely requires fundamental reform. We need to think seriously about how to reconfigure institutions and systems in ways that prevent or mitigate GPG, compensate future people who suffer as a result of GPG, and hold to account relevant agents who pursue GPG.

To illustrate, in previous work, each of us has offered proposals for the kinds of changes that should be considered. One is a global constitutional convention focused on protecting future generations (Gardiner 2014a, 2019). Another is global legal reform that includes making existential threats against the future ('postericide') subject to serious sanctions (McKinnon 2017, 2021). We would also advise taking seriously the idea of an Intergenerational Geoengineering Compensation Fund, such that if geoengineering is pursued by the present generation then those responsible for that pursuit are required to pay into the fund, and future people can make compensation claims if they are damaged by the geoengineering initiated by generations before them. In addition, thought should be given to how to protect nonhuman nature (e.g., by developing international laws against 'ecocide').

Our main point here, however, is not to push specific proposals. It is that conventional proposals often implicitly overestimate – sometimes radically – the capabilities of current institutions for dealing with the challenges associated with SSI, especially for governance across generations. Thus, worries about GPG seem more than reasonable.

The third kind of complacency runs even deeper: some analyses seem to manifest *theoretical complacency*. For instance, some early work suggested that SSI interventions will "benefit everyone"; similarly, some have analyzed SSI as a "global public good". Moreover, both accounts have been used to suggest that SSI escapes many of the usual problems facing international climate policy. Yet such characterizations are usually optimistic at best, and deeply misleading at worst (Gardiner 2013b, 2014; Hourdequin 2018).[30]

Sadly, all three kinds of complacency (political, institutional and theoretical) suggest the lurking presence of moral corruption, and often in ways that raise worries about GPG. For instance, it is predictable that a generational elite tempted by using GPG to protect itself would promote the idea that SSI would be "good for everyone" and easy to govern. Yet a quick reality-check reveals that it is almost certain that SSI (like most other large-scale policies) will have winners and losers, that decisions over it carry with them a tremendous amount of power, and that this is likely to generate conflict. Such issues may not matter much if you are the ones with your hands on the levers of power; still, to ignore or downplay them is to

[30] Fortunately, such claims now seem much less fashionable. For example, most modelers are very clear that they expect winners and losers (e.g., MacMartin et al. 2022).

obscure some of the most fundamental ethical issues at stake in geoengineering. Given all this, another step toward protecting against GPG would be to demand that proposals for developing SSI become much more serious about governance and institutional reform.

5. Objections

5.1. Uncertainty

One scientific objection (to our suggestion that current SSI research encourages GPG) rejects our criticism of the short time-horizons of SSI models, saying that these are unavoidable given the uncertainties involved in climate projections.[31] Specifically, at some point beyond 50-100 years, other uncertainties – such as the evolution of the global economic system or the nature of scientific progress – start to overwhelm the ability of models to project the effects of SSI.[32] Since, it is said, the point of having models is to increase our knowledge of the probability of various outcomes, if they do not do this, there is no good reason to have them.

We understand the basic concern, but remain unconvinced. First, we question whether it is ethically reasonable to proceed with SSI if reasonable long-term projections are not possible. Flying blind in this way seems, on the face of it, to pose huge risks to future generations. It is difficult to imagine that they would approve of the experiment without some level of reassurance as to the longer-term consequences. Surely this, if anything, is a warning sign that GPG is a live threat.

Second, few have tried to model SSI over the much longer term. It may well be very difficult, but a real commitment to avoiding GPG requires at least making a serious attempt. Moreover, some climate models have this temporal reach, and the uncertainties navigated by these models are arguably as great, if not greater, than those facing long term SSI modellers. If climate modellers are at least trying to do this, why not SSI modellers? In addition, as we noted above, the few who have tried have come up with interesting conclusions.

Third, we question the more specific claim that the only point of models is to enable better informed *probability* judgements of various possible outcomes. For one thing, it is already the case that the scenarios used by the IPCC for different

[31] MacMartin et al. 2022 explicitly defend the limited temporal scope of much of the modelling. We address their specific claims in work in progress (Gardiner & McKinnon, in preparation). In order not to overwhelm what is already a long paper, here we consider the issues at a more general level.

[32] Scientists often report that scenario uncertainty becomes the biggest source of uncertainty after 40-50 years or so (e.g., Hawkins 2009, Figure 4). We thank Tom Ackerman and Cecilia Bitz for discussion on this point.

emissions trajectories are informed by models, but do not deliver robust probability assessments and are not designed to do so. For another thing, we think that establishing the extent and range of uncertainty through the use of longer-term models could be extremely useful: it would bring future SSI deployment scenarios within the purview of precautionary approaches, which help to avoid GPG (e.g., Hartzell-Nichols 2012; McKinnon 2019).

5.2. Urgency

A second objection to our account concerns urgency. For instance, some imply that the fast-start focus is appropriate because humanity is so close to breaching the 1.5 and 2.0°C thresholds. Since such breaches threaten climate catastrophe, they suggest, early intervention is necessary even if it comes with extra risks.

We have two basic responses. First, there is a worry about begging the question. The time-constraints associated with responsible development of SSI (e.g., around testing, institutions, etc.) are *already aimed at* determining what kinds of SSI might reasonably be tried and reducing the risks of trying them. Thus, while it is true that humanity may face a "risk-risk" tradeoff (e.g., Parson 2021) or "lesser evil" choice (e.g., Jamieson 1996, Gardiner 2010), it would be a mistake simply to assume that this tradeoff is strongly in favor of a very fast deployment of SSI, especially at this early stage of research and within a governance vacuum. Serious work would need to be done to make that view plausible, which is one reason why a research program is needed.

Our second response is that comparisons with catastrophe can be treacherous. For instance, at first glance it may appear automatically true that SSI would be better than very severe climate change, since the latter is, by definition, genuinely catastrophic. However, such arguments can mislead (e.g., Gardiner 2013b, 2022b). First, if by 'catastrophe', we mean extreme outcomes such as the suffering and death of billions of people or the extinction of humanity, then it is worth emphasizing that SSI is actually only being asked to meet a very low bar of justification: for it can seem that *almost anything* is better than these extremes. Second, being "slightly better than complete catastrophe" is not very impressive or comforting. For instance, meeting the low bar may be easy to achieve, and a characteristic that might be shared with some very unattractive and unethical proposals (e.g., installing a global dictator intent on drastically reducing the human population through genocide). Third, we should not simply *assume* that fast-start SSI passes the low bar, that it is the only policy that would pass it, or that passing it is the threshold we are interested in. Fourth, most strikingly, making the relevant criterion merely the *bare possibility* that fast-start SSI might be better than catastrophic climate change is a clear

mistake. Plausibly, before rushing into SSI on such a thin basis, we should at least consider other strategies within the climate portfolio, including slower-start SSI as well as other radical solutions (e.g., Fragniere and Gardiner 2016). To be clear, our point is *not* that urgency is not an important issue, but that it requires deeper analysis. One should not be too quick to assume that urgency obviously and decisively favors fast-start SSI.

6. Conclusion

Our aim in this paper has been to motivate the idea that the threat of generationally-parochial geoengineering ought to be a core concern of both the ethics of geoengineering and any serious scientific, political or policy discussion. To do this, we explored the concept of GPG, suggested some salient scenarios, and identified early warning signs in the current scientific and policy literature. Within science and policy, the early warning signs include short-time horizons, fast-start focus, and neglect of exit scenarios. When it comes to governance, there is evidence of a status quo bias and of forms of political, institutional, and theoretical complacency that amount to underestimating the task at hand.

Ideally, our discussion will inform development of SSI in ways that help to moderate or even pre-empt GPG. At a minimum, we hope to have done enough to establish that the threat of GPG is sufficiently serious that SSI researchers in all areas should raise the level of alert in their communities, and be on their guard for blind spots, implicit biases, and unnoticed lapses.[33] Moreover, it is encouraging that some first steps for combatting GPG appear straight-forward (e.g., dropping the fast-start focus, exploring a range of scenarios with different start dates).

Nevertheless, in our view the good intentions of researchers are unlikely to be sufficient to protect against GPG. For one thing, often the early warning signs are at the level of norms, assumptions, practices, and shared standards. Good intentions alone do not necessarily control these drivers. Thus, a more robust, and distinctively ethical approach will ultimately be needed.

For another thing, focusing on research alone is inadequate. GPG threatens to impose unjustified risks on future people that involve severe injustices and major violations of legitimacy. Unfortunately, existing institutional architectures are ill-equipped to cope with intergenerational threats. Thus, our identification of GPG and early warning signals will not in itself deliver a shield against intergenerational injustice. Instead, addressing the challenge of GPG is likely to involve serious - and

[33] As we have emphasized, our purpose is not to accuse geoengineering researchers of bad intergenerational behaviour. Indeed, our hope is to make conversations about blame redundant.

perhaps radical – institutional reform. It may also prompt conceptual reform within moral and political philosophy itself. Ensuring that GPG is a core concern in the geoengineering discourse is therefore only an early step on a much longer journey.

Acknowledgements

Earlier versions of this paper were presented at the University of Exeter (2021), the University of Edinburgh (2021), Leibniz Universität Hannover (2022), the Institute for Future Studies in Stockholm (2022), the Hong Kong Catastrophic Risk Centre (2022), Karlsruhe Institute of Technology (2022), Texas A&M University (2023), the University of Auckland (2023), and the University of Canterbury (2023). We are grateful to those audiences, and especially to Gustaf Arrhenius, Krister Bykvist, Doug Campbell, Tim Campbell, Elizabeth Cripps, Heather Douglas, Jim Fleming, Mathias Frisch, Adam Gibbons, Marco Grix, Tim Lenton, Philip Kitcher, Tim Mulgan, Clare Palmer, Martin Peterson, Tanja Rechnitzer, Orri Stefansson, Olle Torpman, and Krushil Watene. We also thank Cecilia Bitz, Tom Ackerman, Arthur Obst, Rob Wood, and two anonymous reviewers. Early work on this paper was supported by the Leverhulme Trust (grant number: H5434400-VP2-2017-014).

References

Armstrong McKay, David I., Arie Staal, Jesse F. Abrams, Ricarda Winkelmann, Boris Sakschewski, Sina Loriani, Ingo Fetzer, Sarah E. Cornell, Johan Rockström, and Timothy M. Lenton. 2022. 'Exceeding 1.5°C Global Warming Could Trigger Multiple Climate Tipping Points'. *Science* 377 (6611): eabn7950.

Belaia, Mariia, Juan Moreno-Cruz, David Keith. 2021. 'Optimal Climate Policy in 3D'. *Climate Change Economics* 12 (3).

Biermann, Frank, Jeroen Oomen, Aarti Gupta, Saleem H. Ali, Ken Conca, Maarten A. Hajer, Prakash Kashwan, et al. 2022. 'Solar Geoengineering: The Case for an International Non-Use Agreement'. *Wiley Interdisciplinary Reviews: Climate Change* 13 (3): e754.

Burns, Wil, and Simon Nicholson. 2017. 'Bioenergy and Carbon Capture with Storage (BECCS): The Prospects and Challenges of an Emerging Climate Policy Response'. *Journal of Environmental Studies and Sciences 2017 7:4* 7 (4): 527–34.

Callies, Daniel Edward. 2019. 'Institutional Legitimacy and Geoengineering Governance'. *Ethics, Policy & Environment* 26 (1): 324–40.

Carr, Wylie and Christopher Preston 2017. 'Skewed Vulnerabilities and Moral Corruption in Global Perspectives on Climate Engineering', *Environmental Values* 26 (2017): 757–777.

———. 2022. *Climate Engineering : A Normative Perspective*. Lexington Books.

Eastham, Sebastian D., Debra K. Weisenstein, David W. Keith, and Steven R.H. Barrett. 2018. 'Quantifying the Impact of Sulfate Geoengineering on Mortality from Air Quality and UV-B Exposure'. *Atmospheric Environment* 187 (August): 424–34.

Fleming, James R. 2010. *Fixing the Sky*. Columbia.

Fragniere, Augustin, and Stephen M. Gardiner. 2016. 'Why Geoengineering Is Not "Plan B"'. *Climate Justice and Geoengineering: Ethics and Policy in the Atmospheric Anthropocene*, no. October: 15–32.

Gardiner, Stephen M. 2001. 'The Real Tragedy of the Commons'. *Philosophy and Public Affairs* 30.4: 387–416.

_____. 2007. 'Is geoengineering the "lesser evil"?' *Environmental Research Web*.

———. 2010. 'Is "Arming the Future" with Geoengineering Really the Lesser Evil? Some Doubts about the Ethics of Intentionally Manipulating the Climate System'. In *Climate Ethics: Essential Readings*, edited by Stephen M Gardiner, Simon Caney, Dale Jamieson, and Henry Shue, 284–314. Oxford: Oxford University Press.

———. 2011a. 'Some Early Ethics of Geoengineering'. *Environmental Values* 20: 163–188.

———. 2011b. *A Perfect Moral Storm*. Oxford: Oxford University Press.

———. 2013a. 'The Desperation Argument for Geoengineering'. *PS: Political Science and Politics* 46 (1): 28–33

———. 2013b. 'Why Geoengineering Is Not a "Global Public Good", and Why It Is Ethically Misleading to Frame It as One'. *Climatic Change* 121 (3): 513–25.

———. 2014a. 'A Call for a Global Constitutional Convention Focused on Future Generations'. *Ethics & International Affairs* 28 (03): 299–315.

———. 2014b. 'Why 'Global Public Good' Is a Treacherous Term, Especially for Geoengineering'. *Climatic Change* 123 (2): 101–06, https://www.proquest.com/docview/1504829455

———. 2017. 'Climate Ethics in a Dark and Dangerous Time'. *Ethics* 127: 430-465.

———. 2019. 'Motivating (or Baby-Stepping Toward) a Global Constitutional Convention for Future Generations'. *Environmental Ethics* 41 (3): 199–220.

———. 2020. 'Ethics and Geoengineering: An Overview'. In *Global Changes: Ethics, Politics and Environment in the Contemporary Technological World*, edited by Luca Valera and Juan Carlos Castilla, 69–78. Ethics of Science and Technology Assessment. Cham: Springer International Publishing.

———. 2022a. "Is the Paris Climate Agreement Another Dangerous Illusion?". *Norwegian Academy of Science and Letters Yearbook* 2021. Oslo: Novus.

———. 2022b. 'The Ethical Risks of an Intergenerational World Climate Bank (as Opposed to a Climate Justice World Bank)'. IFFS Working Paper.

Gardiner, Stephen M., Catriona McKinnon and Augustin Fragnière. 2021. *The Ethics of "Geoengineering" the Climate*. Routledge.

Gonzalez-Ricoy, Inigo and Axel Gosseries, eds. 2016. *Institutions for Future Generations*. Oxford University Press.

Hale, Benjamin. 2012. 'The world that would have been: Moral hazard arguments against geoengineering.' In C.

Hamilton, Clive. 2013. *Earthmasters: The Dawn of the Age of Climate Engineering*. Yale University Press.

Hartzell-Nichols, Lauren. 2012. 'Precaution and Solar Radiation Management'. *Ethics, Policy and Environment* 15 (2): 158–71.

Horton, Joshua, and David Keith. 2016. 'Solar Geoengineering and Obligations to the Global Poor'. In *Climate Justice and Geoengineering: Ethics and Policy in the Atmospheric Anthropocene*, edited by Christopher J Preston, 79–92. Rowman and Littlefield International.

Hourdequin, Marion. 2018. 'Climate Change, Climate Engineering, and the 'Global Poor': What Does Justice Require?', *Ethics, Policy & Environment* 21(3): 270-288.

Hulme, Mike. 2012. 'Climate Change: Climate Engineering through Stratospheric Aerosol Injection'. *Progress in Physical Geography* 36 (5): 694–705.

Ibbott, Samantha. 2021. 'Solar Geoengineering Not a "Sensible Rescue Plan", Say Scientists | Imperial News | Imperial College London'. 2021.

International Energy Agency (IEA). 'Global Energy Review: CO2 Emissions in 2021'.

IPCC. 2018. 'Global Warming of 1.5 ºC — Special Report'. https://www.ipcc.ch/sr15/.

Jamieson, Dale. 1996. 'Ethics and Intentional Climate Change'. *Climatic Change* 33 (3): 323–36.

Kashwan, Prakash, Frank Biermann, Aarti Gupta, and Chukwumerije Okereke. 2020. 'Planetary Justice: Prioritizing the Poor in Earth System Governance'. *Earth System Governance* 6 (December): 100075.

Keith, David W. 2010. 'Photophoretic levitation of engineered aerosols for geoengineering.' *Proceedings of the National Academy of Sciences of the United States of America, 107,* 16,428–16,431.

Keith, David W, and Douglas G MacMartin. 2015. 'A Temporary, Moderate and Responsive Scenario for Solar Geoengineering'. *Nature Climate Change* 5 (3): 201–6.

Kravitz, Ben. 2020. 'Ten Years of GeoMIP | Harvard's Solar Geoengineering Research Program'. 2020. https://geoengineering.environment.harvard.edu/blog/ten-years-geomip.

Kravitz, Ben, and et al. 2014. 'Explicit Feedback and the Management of Uncertainty in Meeting Climate Objectives with Solar Geoengineering Strategic Incentives for Climate Geoengineering Coalitions to Exclude Broad Participation Explicit Feedback and the Management of Uncertainty in Meeting Climate Objectives with Solar Geoengineering'. *Environmental Research Letters* 9: 44006–13.

Kravitz, Ben, Alan Robock, Olivier Boucher, Hauke Schmidt, Karl E. Taylor, Georgiy Stenchikov, and Michael Schulz. 2011. 'The Geoengineering Model Intercomparison Project (GeoMIP)'. *Atmospheric Science Letters* 12 (2): 162–67.

Lenferna, Georges Alexandre, Rick D. Russotto, Amanda Tan, Stephen M. Gardiner, and Thomas P. Ackerman. 2017. 'Relevant Climate Response Tests for Stratospheric Aerosol Injection: A Combined Ethical and Scientific Analysis'. *Earth's Future* 5 (6): 577–91.

Lenton, Timothy M, Johan Rockström, Owen Gaffney, Stefan Rahmstorf, Katherine Richardson, Will Steffen, and Hans Joachim Schellnhuber. 2019. 'Climate Tipping Points — Too Risky to Bet Against'. *Nature* 575 (7784): 592–95.

MacMartin, D. G., D. Visioni, B. Kravitz, J. H. Richter, T. Felgenhauer, W. R. Lee, D. R. Morrow, E. A. Parson, and M. Sugiyama. 2022. 'Scenarios for Modeling Solar Radiation Modification'. *Proceedings of the National Academy of Sciences* 119 (33): e2202230119.

Malik, Abdul, Pierre Nowack, Joanna Haigh, Long Cao, Luqman Atique and Yves Plancherel. 2020. 'Tropical Pacific climate variability under solar geoengineering'. *Atmospheric Chemistry and Physics* 20: 15461–15485. https://doi.org/10.5194/acp-20-15461-2020

Maltais, Aaron, and Catriona McKinnon. 2015. *The Ethics of Climate Governance*. London: Rowman & Littlefield International, Ltd.

McCusker, K. E., D. S. Battisti, and C. M. Bitz. 2015. 'Inability of stratospheric sulfate aerosol injections to preserve the West Antarctic Ice Sheet'. *Geophysical Research Letters* 42, doi:10.1002/2015GL064314.

McKinnon, Catriona. 2017. 'Endangering Humanity: An International Crime?', *Canadian Journal of Philosophy* 47 (2-3). pp. 395–415. s

———. 2019. 'Sleepwalking into Lock-in? Avoiding Wrongs to Future People in the Governance of Solar Radiation Management Research'. *Environmental Politics* 28 (3).

———. 2020. 'The Panglossian Politics of the Geoclique'. *Critical Review of International Social and Political Philosophy* 23 (5): 584–99.

———. 2021. 'Postericide and Intergenerational Ethics'. In S. Gardiner, ed., *The Oxford Handbook of Intergenerational Ethics*, Oxford University Press.

McLaren, Duncan, and Olaf Corry. 2021a. 'The Politics and Governance of Research into Solar Geoengineering'. *WIREs Climate Change* 12 (3): e707.

———. 2021b. 'The Politics and Governance of Research into Solar Geoengineering'. *Wiley Interdisciplinary Reviews: Climate Change* 12 (3).

Moellendorf, Darrel. 2013. *The Moral Challenge of Dangerous Climate Change*. New York: Cambridge University Press.

Moreno-Cruz, Juan and Keith, David. 2013. "Climate policy under uncertainty: a case for solar geoengineering." *Climatic Change*, 121(3): 431–444.

Morrow, David and Toby Svoboda. 2016. 'Geoengineering and Non-Ideal Theory'. *Public Affairs Quarterly* 30 (1): 85–104.

Morrow, David R. 2020. 'A Mission-Driven Research Program on Solar Geoengineering Could Promote Justice and Legitimacy'. *Critical Review of International Social and Political Philosophy* 23 (5): 618–40.

Morrow, David R, Robert E Kopp, and Michael Oppenheimer. 2013. 'Political Legitimacy in Decisions about Experiments in Solar Radiation Management'. In

Climate Change Geoengineering: Philosophical Perspectives, Legal Issues, and Governance Frameworks, edited by Wil C G Burns and Andrew Strauss, 146–67. Cambridge: Cambridge University Press.

National Academies of Sciences, Engineering, and Medicine ('NASEM'). 2021. 'Reflecting Sunlight: Recommendations for Solar Geoengineering Research and Research Governance'. *Reflecting Sunlight*. National Academies Press.

Nature Editorial. 2021. 'Give Research into Solar Geoengineering a Chance'. *Nature* 593 (7858): 167.

Pamplany, Augustine, Bert Gordijn, and Patrick Brereton. 2020. 'The Ethics of Geoengineering: A Literature Review'. Science and Engineering Ethics 26 (6): 3069-3119.

Parker, Andy, and Peter J Irvine. 2018. 'The Risk of Termination Shock From Solar Geoengineering'. *Earth's Future*, 1–12.

Parker, Andy, Joshua Horton, and David Keith. 2018. 'Stopping Solar Geoengineering Through Technical Means: A Preliminary Assessment of Counter-geoengineering'. *Earth's Future, 6*, 1058–1065.

Parson. Edward A. 2021. 'Geoengineering: Symmetric precaution', *Science* 374 (6569): 795.

Parson, Edward A., and Jesse L. Reynolds. 2021. 'Solar Geoengineering: Scenarios of Future Governance Challenges'. *Futures* 133 (October): 102806.

Patrick, Stewart. 2022. *Reflecting Sunlight to Reduce Climate Risk*. Council on Foreign Relations, Special Report 93.

Pierrehumbert, Raymond. 2019. 'There Is No Plan B for Dealing with the Climate Crisis'. *Bulletin of the Atomic Scientists* 75 (5): 215–21.

Preston, Christopher J. 2012. *Engineering the Climate: The Ethics of Solar Radiation Management*. London: Rowman and Littlefield.

———. 2013. 'Ethics and Geoengineering: Reviewing the Moral Issues Raised by Solar Radiation Management and Carbon Dioxide Removal: Ethics & Geoengineering'. *Wiley Interdisciplinary Reviews: Climate Change* 4 (1): 23–37.

———. 2016. *Climate Justice and Geoengineering : Ethics and Policy in the Atmospheric Anthropocene*. London: Rowman & Littlefield International, Ltd.

Rabitz, Florian. 2018. 'Governing the Termination Problem in Solar Radiation Management'. *Environmental Politics*, September, 1–21.

Rogelj, Joeri, Michel den Elzen, Niklas Höhne, Taryn Fransen, Hanna Fekete, Harald Winkler, Roberto Schaeffer, Fu Sha, Keywan Riahi, and Malte Meinshausen. 2016. 'Paris Agreement Climate Proposals Need a Boost to Keep Warming Well below 2 °C'. *Nature* 534 (7609): 631–39.

Shue, Henry. 2017. 'Climate Dreaming'. *Journal of Human Rights and the Environment* 8 (2): 203–216

Stephens, Jennie C, Prakash Kashwan, Duncan Mclaren, and Kevin Surprise. 2021. 'The Dangers of Mainstreaming Solar Geoengineering: A Critique of the National Academies Report'.

Svoboda, Toby. 2012. 'The Ethics of Geoengineering: Moral Considerability and the Convergence Hypothesis'. *Journal of Applied Philosophy* 29 (3): 243–56.

The Economist. 2021. 'Geoengineering Is Conspicuously Absent from the IPCC's Report | The Economist'. *The Economist*, 14 August 2021.

UNEP. 2019. 'Emissions Gap Report 2019'. https://www.unep.org/resources/emissions-gap-report-2019.

———. 2020. *Emissions Gap Report 2020*. Nairobi: United Nations Environment Programme. https://www.unep.org/emissions-gap-report-2020.

———. 2021. 'Emissions Gap Report 2021'. https://www.unep.org/resources/emissions-gap-report-2021.

———. 2022. 'Emissions Gap Report 2022'. https://www.unep.org/resources/emissions-gap-report-2022.

United Nations. 1992. 'United Nations Framework Convention on Climate Change'. *UNFCCC*.

———. 2009. 'Copenhagen Accord'.

———. 2015. 'Paris Agreement'. *United Nations*.

Vukajlovic, Johvanna et al 2021. 'Diamond-doped Silica Aerogel for Solar Geoengineering', *Diamond & Related Materials* 117 (2021) 108474.

Whyte, Kyle Powys. 2018. 'Indigeneity in Geoengineering Discourses: Some Considerations'. *Ethics, Policy & Environment* 21 (3): 289–307.

Clare Heyward[1] & Edward Page[2]

Rectifying Secondary Climatic Injustices

Due to faulty planning or unforeseeable contingencies, policies undertaken to manage climate change may succeed in reducing one source of disruption in peoples' lives only to introduce a new source of disruption. Where these disruptions would be intolerable without further intervention to ameliorate them, we can say that a 'secondary climatic injustice' has arisen. Secondary climatic injustices can usefully be distinguished from 'primary climatic injustices', which concern unjustified disruptions of peoples' lives that arise due to the absence of policies designed to manage climate change. In this paper, we show how secondary climatic injustices arise from multiple pathways of policymaking and then set out an account of how these injustices can be rectified by compensating the victims so that, even if they do bear some additional costs, they share the costs of tackling climate change equitably with other users of the climate system. This basic level of compensation, we argue, may be enhanced if one or both of two exacerbating features arise on the part of the policymakers who cause a secondary injustice. These are (i) how avoidable the secondary injustice was from the policymaker's perspective, and (ii) how excusably ignorant the policymaker was for not selecting the most just policy.

[1] Clare Heyward UiT: The Arctic University of Norway. jennifer.c.heyward@uit.no

[2] Edward Page, University of Warwick e.a.page@warwick.ac.uk

1. Introduction

It has long been known that climate change will have severe impacts that, without intervention, will lead to widespread loss and damage (IPCC 2014: 14; IPCC 2023: 24–27). In most cases, those who will be most adversely affected will be the poorest and, historically speaking, least responsible for emergence of climate change (IPCC 2023: 5). The seriousness of these loses and damages has led to the adoption of a series of international and domestic policies of *mitigation* (actions to limit changes in climate by limiting increases in atmospheric greenhouse gas concentrations) and *adaptation* (actions to ensure that any climatic changes that do occur do not result in harmful effects on human well-being).[3] Normative theorists have made a significant contribution to mitigation and adaptation policy by providing rigorous critique of the arguments and justifications offered for alternative policies. A common point of agreement, however, is that significant mitigation and adaptation is required, funded primarily by developed states, as a matter of global and intergenerational justice (Caney 2005, 2011a,b; Moellendorf 2014; Vanderheiden 2008; Wallimann-Helmer 2015).

More recently, the focus of climate scholars has expanded beyond "mitigation" and "adaptation" to include up to five separate policy types (*emissions reduction*, *sink enhancement*, *solar radiation management*, *adaptation*, and *compensation*) that reside on a continuum from prevention to rectification.[4] The normative literature on climate change, to the extent it has evaluated and recommended action under this expanded list of policies, has typically emphasized the way that unmanaged climate change will lead to disruptions of lives in developed and developing states (see Eckersley 2016; Page and Heyward 2017; Wallimann-Helmer 2015). Where such disruptions cannot be justified – for example, where they arise for populations who are the least able to cope while also having played the smallest role in driving climate change – they can usefully be seen as 'primary climate injustices' since they arise from changes in climate that arise *despite* attempts to manage climate change.

There is, however, a further source of injustice that has been largely neglected by climate change scholars: adverse effects of climate change that arise *from* measures implemented to manage climate change. It is important to note that not all of the undesired impacts of climate policy will be unjust but, so long as they surpass a threshold of a harm and cannot be justified to those that suffer from them, they should, we argue, be seen as 'secondary climatic injustices.'[5] Secondary injustices arise from imperfections, foreseeable or otherwise, in the design or operationaliza-

[3] In the paper, we refer interchangeably to 'policies', 'measures' and 'responses.'

[4] For an early exploration of this expanded taxonomy of climate policy, see (Heyward 2013).

[5] Henceforth we shall often refer simply to "primary injustices" and "secondary injustices".

tion of climate policies that do not interfere with the realisation of the objectives specified in the case for their adoption. They may also arise from unpredicted, or unpredictable, responses of human and natural systems to policies during or after implementation. Since secondary injustices are human originating, highly debilitating, and would not have arisen in a world free of climate change, there is little reason why they should not be given an equally prominent role in normative evaluations of climate change as primary injustices.[6]

In what follows, we develop, in Section 2, the concept of secondary climatic injustice and give some examples and a rough typology. Section 3 then argues that it is appropriate to rectify secondary injustices, but not necessarily in the same way as primary climatic injustices. Unlike primary injustice, the just rectification of secondary injustice need not always involve making victims fully ″whole again." It can, instead, aim at a form of constrained compensation that restores a "rough justice" between the affected parties. This section also sketches two normative criteria that play a role in modifying the exact amount of compensation a victim of a secondary injustice should receive. Section 4 concludes.

2. The surprising familiarity of secondary climatic injustices

Adverse impacts of policies designed to tackle climate change, although under theorised in the normative literature, are well documented by academics and policymakers concerned with the complexity and contingency of climate policymaking. Barnett and O'Neill (2010: 221) explore how certain adaptation measures, designed to reduce the vulnerability of some groups to climate change, may result in 'an increase in vulnerability of other systems, sectors or social groups.' Sidi (2012: 349) explores how certain mitigation measures, designed to reduce the greenhouse gas emissions of some sector or social group, may lead to equal or greater emissions by other sectors or social groups. Developing country negotiators have also referred to the dangers of bringing about undesirable impacts through climate policy in their submissions to the UN climate negotiations. Typically, their concern has been that policies designed to reduce greenhouse gas emissions in developed states ought not disrupt the economies of developing states without compensation being offered to offset these disruptions (Chan 2016).

At the same time, acknowledgment of the adverse impacts of climate policies has taken place in the broader context of great skepticism as to whether current commitments across all types of policy will lead to *any* meaningful checks on climate

[6] Our claim here is based largely on similarity of moral treatment for morally similar phenomena.

change. So, the focus of the climate change community as a whole has largely been on how to encourage more ambitious climate policies and not on the undesired impacts of those few policies that have made it through to implementation. This is troubling for two reasons. First, where the undesired impacts of existing policies and measures can reasonably be seen as imposing an unfair burden on already impoverished populations, we are presented with the puzzle of whether these effects should trigger additional action, such as compensation, in the name of climate justice. This is the puzzle of past secondary climate injustices. Second, even if undesired climate impacts are relatively small at present, given the limited scope of the climate response, they will undoubtedly increase as more ambitious climate policies are undertaken in the name of climate injustice increasingly shape the lives of all populations. This is the puzzle of future secondary climate injustices. In what remains of this section, we explain how secondary injustice is an integral feature of climate policymaking as a precursor to a sketch of a normative account of how such injustice should be rectified.

2.1. Moving the side effects to the centre

Although adverse impacts of climate policies have been addressed in the literature, there has been some doubt as to the features of the policy, or the wider environment, that are most likely to trigger such impacts. For example, a common objection to the increased use of biofuels, an emissions reduction measure, is that the land needed to produce biofuels may need to be taken from the poorest and most marginalised groups without their consent (Nuffield Council on Bioethics 2011). But there has been little discussion on how this injustice, if it is indeed an injustice, might be rectified. Solar radiation modification (also called solar radiation management) technologies (SRMs) may also bring about adverse side-effects, such as changing precipitation patterns, that could be seen as unjust if they increased disadvantage in vulnerable populations (Shepherd et al 2009; NAS 2014b). Implementing certain adaptation measures, such as coastal migration, may also prove beneficial to one group while imposing intolerable costs on other groups. Finally, early warning systems, compensation programmes, and climate insurance schemes – which are designed to manage losses and damages that remain after all other policies have been attempted – may exacerbate existing vulnerabilities if they succeed in addressing one group's loss by shifting losses to another group for example by prolonging environmentally destructive behaviours. The undesired disruptions in all these examples would have been deemed unjust if they resulted from unmanaged climate change and so, we shall argue, should be seen as unjust even though they arose from "good faith" efforts to manage climate change.

The adverse impacts of climate policies have appeared in scientific and negotiation texts under the rubric of "side effects" [7] and "impacts of implementing response measures";[8] and they have also appeared sporadically in the normative literature.[9] However, in both cases, undesired impacts are discussed in the context of the problem of whether it is *permissible* to implement a particular response measure and not as a puzzle of how the undesirable impacts of permissible climate responses should be managed. As a result, the hardships experienced by the victims are set aside so long as the policy in question is seemed a permissible method of tackling a primary climate justice *all things considered*. Simon Caney (Caney 2011a:172), for example, has argued that policies of mitigation or adaptation that violate human rights in order to counteract the adverse impacts of climate change should not be undertaken;[10] and Eriksen et al. (2011) argue that adaptation policies must not shift risks onto (or reduce the adaptive capacity of) future generations since this would violate a principle of "sustainable adaptation" (Eriksen et al. 2011). Secondary injustice questions relating to SRMs, such as whether, as David Morrow puts it, one may "cause a flood to stop a fire" (Morrow 2014:123), have also focused on the question of whether it is permissible to cause some collateral damage in the drive to reduce the causes of primary climate injustice without addressing what should be done about secondary injustice (the loss and damage associated with 'the flood').

The lack of discussion about how to respond to secondary injustices is troubling for two reasons. First, the creation of undesirable effects is an inescapable part of climate policymaking and may arise from even the most well-designed policy. Similarly, even the best designed policy may create impacts that it would be wrong to leave to continue to blight the lives of those initially affected. One issue is that both human beings and the institutions they create are inherently fallible. Accordingly, "honest mistakes" and "policy failures" are an integral feature of all forms of public policy. Climate change is also uniquely complex. This makes it unavoidable that mistakes, failures, and adverse side effects will occur. Climate change has, in

[7] Side effects are the effects of policies that did not feature in the justification for undertaking the policy. They are generally, but not always, unanticipated, unintended, and unexpected. Chan (2016: 228) usefully distinguishes between 'first order effects' (the unprevented and unpreventable negative effects of climate change) and 'second order effects' (the negative effects of attempts to control climate change on the economic development of affected, particularly developing, states).

[8] The UNFCCC obliges parties to give 'full consideration, in the implementation of the commitments of the Convention, the specific needs and concerns of developing country Parties arising from the impact of the implementation of response measures. When addressing climate change concerns, the Kyoto Protocol commits Parties to strive to minimize adverse economic, social and environmental impacts on other Parties, especially developing country Parties' (see http://unfccc.int/cooperation_support/response_measures/items/4908.php).

[9] See e.g. Caney 2011a; Eriksen et al. 2011; Morrow 2014:123.

[10] Caney suggests that victims of such transgressions are due compensation (2011a: 171) but does not elaborate.

this way, been described as a "wicked" problem (see Hulme 2009; Incropera 2016; Ney and Verweij 2015). Even if this complexity and "wickedness" is understood by policymakers, this will not prevent mistakes being made which could then lead to secondary injustices.

Second, concerted action on climate change should have begun many decades ago and/or should have been implemented at a much faster and larger scale. Because it was not, losses and damages are already occurring due to climate change - and it is no longer possible to prevent all future losses. In these less-than-ideal circumstances, trade-offs will have to be made. As we saw above, some theorists have cautioned that some secondary injustices might be sufficiently grave to render impermissible the connected policy. However, we must also allow for the possibility that, even if a secondary injustice is not sufficiently grave to do this, the problem remains as to how the impaired condition of the victims should be rectified. For example, policies designed to reduce greenhouse gas emissions in developed states may save the developing world from severe climate impacts in the future. It does not follow, however, that the negative effects of these policies on businesses in developing states should be left to blight those affected. The policymakers in such situations seem to face a tragic choice. In such cases, creating some injustice is unavoidable from their perspective even though this does not involve wrongdoing provided they choose the best response available, i.e. that which reduces primary climate injustices and attempts to avoid secondary ones. In what follows, we focus our concern on this important category of secondary climatic injustices, namely, disruptions in peoples' lives that it would be wrong to leave uncorrected even though they owe their existence to climate policies that are permissibly undertaken.

2.2. Three key pathways to secondary injustice

Having established that side effects are an integral feature of climate policymaking, the task remains to clarify the specific categories of secondary injustice that may arise from side effects. We argue that there are three such categories each linked to a separate pathway through which permissible policies may generate secondary injustices.

2.2.1. Unfair distribution of implementation burden (category 1)

Implementing any climate policy will require some agents to bear costs over and above those they would have borne otherwise. Some of these burdens will be borne by project developers, others will be borne by citizens through general taxation, and others will be borne by those who 'pay' for the policy in some other way. The term

"burdens" should be understood broadly as including the economic and non-economic costs of implementing a climate policy and will include everything from the policy's operational financial cost to the adjustments in lifestyle, cultural and spiritual practices required by the policy for it to be a success. If a response to climate change is successful in its own terms but imposes disproportionate and unfair costs on those who pay for the policy, a secondary injustice may arise. Often, the risk of an unfair set of burdens of implementation would be seen by policymakers in advance and may even make the policy impermissible if other alternatives are available which lack this unfairness. Even if this category of secondary injustice is rarely instantiated, it is at least possible that a climate policy has sufficient merit that it is permissible, all things considered, even if it introduces significant unfairness in the distribution of burdens of implementation.

Different types of policy response will be vulnerable to specific instances of implementation burden injustice. An adaptation policy, for example, might benefit a particular disadvantaged group as planned but to the detriment of another group that disproportionately pays for the scheme despite the latter having become similarly worse off or marginalized since the adoption of the policy. A further set of examples relates to land-use change policies that reduce emissions (or enhance sinks) of greenhouse gas at the cost of raising the risk of conflicts over tenure and access to land. The concern here is that indigenous peoples may pay a disproportionate price for the success of the policy in terms of land dispossession. A similar, related, issue is that of "justice in siting" (Hunold and Young 1998). Although the disadvantaged may not be prevented from accessing or using a piece of land, they are instead forced to cope with having hazardous, or otherwise unappealing, industrial projects built near their homes or near sites of special significance to their communities. Controversies about the siting of carbon capture and storage (CCS) plants (e.g. Fischer 2014), show that various emissions reductions schemes may be unwelcome to the local population, and, depending on the conditions of that population, constitute a case of unjust siting. The same may be the case of sink enhancement projects which involve large-scale industrial engineering, such as direct air capture projects and enhanced weathering projects which require large quantities of minerals to be mined (Shepherd et al 2009: 14).

2.2.2. Displacement of climate impacts (category 2)

A second category of secondary injustice arises from policies that successfully reduce one type of climate impact only to cause (or exacerbating) a different type of impact. Emissions-reduction and sink enhancement policies, for example, may require large-scale land-use changes to reduce global warming (such as the planting

of fuel crops and trees in "plantation style") only to reshape local biodiversity and agriculture. If these side effects are sufficiently severe, they may constitute a secondary injustice by damaging the livelihoods of local populations who rely on land and marine species for subsistence. SRM technologies counteract global warming by reflecting the sun's energy back into space but some of them (particularly aerosol injection) come with the risk that, if they fail or are intentionally terminated, rapid climate change could result that causes severe impacts (Shepherd et al 2009: 26; IPCC 2021: 37). Some of the adverse effects risked through by the policies describe above may seem so grave that the initial implementation of the relevant technologies should have been seen as impermissible. In other cases, the matter may be more finely balanced and so raise issues of secondary injustice if the adoption of the technology reduced global warming but increased ocean acidification and sea level rises.

Finally, some adaptation measures may displace environmental impacts of climate change from one group to another even if the combined impact is reduced. The IPCC cautions that some schemes designed to combat coastal erosion may result in erosion taking place further down the coastline (Noble and Huq 2014: 858). Similarly, diverting a river, or drawing extra water from it, may be a good form of adaptation for a community at risk of draught or desertification, but one that leaves downstream communities short of water, thereby increasing their vulnerability to the very effects the upstream community sought to avoid. Whereas the balancing of the interests of different communities may have been permissible when the policy was adopted, the policy in implementation has led to displaced impacts seen as intolerable for those affected.

2.2.3. Undermining the climate effort (category 3)

Even where a policy brings about the desired change in the selected dimension of climate change, a secondary injustice may arise if the policy undermines the contribution or commitment of the affected populations to other climate policies. There are three pathways through which this injustice might be brought about: *incentives reduction*, *perpetuating vulnerability*, and *self-defeat*. In each case, though a policy was permissibly enacted and successful in its own terms, a secondary injustice is generated though the undermining of another part of (or kind of) the climate response.
Incentives reduction. Climate policies may sometimes succeed in their own terms but at the cost of reducing incentives to engage in other kinds of climate policy. Reducing incentives to cut greenhouse gas emissions in a certain way, for example, may discourage the take up of alternative emissions reduction measures if the original policy unexpectedly fails or is discontinued for some other reason. This would result in a greater need for CDRs, SRMs or adaptation measures if loss and damage

is to be avoided. The 'moral hazard effect' has been one of the most frequently raised concerns about sink enhancement and SRM technologies but the incentives reduction problem arises across all policy types. Barnett and O'Neill, for example, warn that mismanaged adaptation policies could reduce incentives to adapt 'by encouraging unnecessary dependence on others, stimulating rent-seeking behaviour, or penalising early actors' (2010: 212). In fact, virtually any climate policy if poorly implemented could discourage the take up of alternative policies and thereby impose additional costs on others.

Perpetuation of vulnerability. Climate policies may inculcate a false sense of security amongst agents to the effect that the policy has removed a climate threat when the threat remains present. The IPCC gives an example of people and developers moving to areas where coastal protections are being built (Noble and Huq 2014:858). Should these protections turn out to be inadequate (e.g. because the climate change-caused storm surges are higher than predicted or demographic changes render the defences inadequate to protect a growing population) then the policy may succeed in narrow terms only at the cost of increasing vulnerability of some populations.

Self-defeat. A response measure, though permissible at the time of adoption, may unexpectedly unravel due to the emergence of hidden costs. In such cases, the policymakers may have engaged in sufficient deliberation and planning to render the policy permissible, but circumstances change so much that the policy wastes valuable time and consumes valuable resources in a way that disrupts lives without any tangible reward. A good example here is a preventative response that reduces emissions as planned in one sector only at the cost of using more energy in its implementation than was reduced. The idea is that once the hidden defects of some policies are fully understood, these policies cannot reasonably be seen as injustice-free despite their objectives, narrowly defined, being achieved. One example would be policies encouraging greater use of air conditioning to counteract heat waves that require greater fossil fuel consumption which may then make mitigation more costly harder and also crowd out investment in less carbon intense adaptation policies and behaviours.

To conclude this analysis of secondary injustice pathways, two points of clarification should be added. First, secondary injustices are not *policy failures* in the sense that they are the result of a policy failing to meet its stated objectives. It is not the failure of the policy that explains how a secondary injustice arises but rather the imposition of intolerable burdens in pursuit of the success of this policy. Policies generate secondary injustices when their second order effects are intolerable and

not because they failed in terms of their objective to bring about the desired first order effects. The failure of an otherwise permissible and successful emissions reduction programme to achieve the stated emissions cuts, for example, does not generate secondary injustice so much as leave intact a primary injustice. Second, our aim here has been to highlight the *possibility* of secondary injustices. Although they appear impossible to avoid entirely, it is by no means clear how often they will occur or how severe they will be in comparison to primary injustices. As we have seen, history, plus the complexity of the problem of climate change, give us reason to think that there will be extensive secondary injustice and it is a mark of a comprehensive theory of climate change justice that we prepare for this.

3. Rectifying climate injustices

In this section, we argue that an attractive approach to secondary injustice is to treat it as a problem of constrained (or 'rough') compensation whereas primary injustice is naturally treated as a problem of full and immediate compensation. As we saw above, both primary and secondary climate injustice arise from disruptions linked to anthropogenic climate change and both involve lives being disrupted through no fault of their own in a way that will not be repaired without further intervention. Both injustices will also be most disrupting for communities that have done least to cause climate change, benefited least from the economic practices that drive climate change, and have the least ability to respond.[11] There are, however, important normative differences between these two sources of injustice which indicate that they have different solutions.

The basic case for addressing primary climatic injustice is that agents who suffer a significant drop in well-being due to the wrongful actions of others ought to be compensated because they have suffered an *unjustifiable disruption* to their lives. These disruptions compromise the valued ends that people purpose and/or the means that people use to shape and pursue these ends (Page and Heyward 2016). Some primary injustices may be sufficiently trivial that they do not pass a *de minimis* test that the adverse disruption must be of sufficient moral importance to trigger compensation. But, if this test is met, victims are due *full* and *immediate* compensation so that their relationship with the means at their disposal, as well as the valued ends pursued with these means, is restored as closely as possible to what it was prior

[11] In pure economic terms, the cost of secondary injustice may be highest in developed states due to the higher combined value of disrupted assets located within these states, but we focus here on the secondary injustices arising in developing states since the citizens and institutions in these states can sometimes have less capacity to respond through domestic action. This is effectively to apply the 'common but differentiated responsibilities' principle to secondary climate injustice as it has often been applied to primary climate injustice.

to their experience of the injustice (Goodin, 1995:484-5).[12] Receiving such compensation makes the victim's life 'whole again' in the sense that (to the greatest extent possible), it is as if the unjustifiable interruption of the victim's life had never happened.[13]

Secondary injustices, by contrast, arise from *justifiable* disruptions in peoples' lives associated with attempts to counter the primary injustices of climate change. This crucial difference raises the possibility that it may not always be appropriate to make the victims of secondary injustice "whole again" to rectify the injustice between these victims and those agents responsible for their impaired state. Why might this be the case? So long as these policies were permissibly undertaken, the "victims" were not wrongfully treated by the policymakers and so may be reasonably required to bear *some* of the cost of adverse secondary effects along with the policymakers. To put it into crude terms, policymakers may say to the victims something like the following:

'We can see that your life has been disrupted and you have lost out in the way the policy was designed or implemented. You did nothing to deserve such an interruption. However, had we not taken this course of action, we would have violated our responsibility to all those who suffer unjust disruption from climate change and these people are similarly undeserving. Indeed, you were included in this population, seen impersonally, since we undertook the action for everyone. So, we haven't wronged you so long as you are not singled out for an unfair burden of these additional costs."

The idea is that climate policies that but cause undesirable effects may be justifiable *all things considered* so long as these policymakers set aside sufficient compensation to those unfairly affected. But what would be *sufficient*? In our view, compensation for secondary injustice is sufficient not when it is indexed to a situation where the disruption at the heart of the injustice had never happened but rather to a situation where the unplanned disruption of an otherwise successful and permissible policy is equitably shared amongst victims and non-victims. We might call this the 'fair burdens baseline.' Secondary injustices are corrected, on this view, when no

[12] This is a rectificatory approach to secondary injustice since it seeks to make it the case that the injustice between the perpetrator, in this case the policymaker, and the victim, who suffers from secondary injustice due to the actions of the policymaker, had never happened. It can be usefully contrasted with a "distributive approach" that would compensate for the injustice only if, and to the extent that, it departs form a preferred ideal of distribution. We shall not discuss the latter here except to note that rectificatory approaches to secondary climate injustices are compatible with viewing other parts of the climate problem in distributive terms. We should also to note that the rectificatory approach defended in the text is somewhat pluralist in incorporating elements of distributive justice in the way initial compensation for secondary injustice is measured.

[13] On the idea of correcting an injustice as a matter of making it as if wrongful transactions had never happened, see Ripstein (2007: 1993) and Gardner (2012: 28–31).

one can complain that they have been singled out for unfair treatment in the bearing of the secondary injustice as they would be if they were the initial victim, and the costs were left for them and no-one else. If this fairness in burden sharing is achieved, it is as if the secondary injustice had never happened since no agent will be paying an unacceptable price for the implementation of a *just* climate policy. For example, suppose that policymakers and their expert advisers have a reasonable belief that a sink enhancement project should be located in a rural area. The required industrial engineering will require large quantities of minerals to be mined and this will have an adverse impact on the local community which is already suffering from significant disadvantage. The argument here is that the local population should be compensated for the disruption incurred so that they bear a roughly equal burden of the climate response even if this does not return to them to the same level of satisfaction or wellbeing as they would have enjoyed had the installation been sited elsewhere.[14]

3.1. Modifying compensation

Are there circumstances where the victims of secondary injustices can demand more than the minimum level of compensation as described above? Here we propose *two* factors that boost the amount of compensation that victims might reasonably demand of the policymakers responsible for a secondary injustice.[15] The

[14] It might be objected that remedying secondary injustice in the way suggested raises the potential for an endless regress since, if climate policies cause a secondary injustice, then responding to this injustice by restoring fairness in the sharing of burdens of adverse policy side effects could be expected to create further injustices which then would require remedy. Even the constrained compensation associated with the 'fair burdens baseline' can be seen as a form of policy, liable to defects, that will create further injustice. Wouldn't it be better just to 'let it be'? Consider an SRM project enacted by P, in good faith, which causes a secondary injustice to V through increasing precipitation and flood events in V's region. V is then compensated by P by cash payments and new flood defences so P no longer bears a disproportionate burden of the secondary injustice caused by the SRM project. If this new policy turns out to be maladaptive in any way (perhaps the flood defences induce a boom in migration to the area which causes a collapse of healthcare) then a tertiary injustice seems to arise for a new set of victims that requires even more compensation and so on and so on. But this certainly overstates the problem. First, the disruption in question, to be a secondary injustice, cannot have arisen 'but for' the impugned policy. This will become much harder to establish over time in the same way that any effect is harder to trace back to its causes in the more distant past. Second, the disruption in question, to be a secondary injustice, must meet the de minimis challenge. This will also become much harder to establish over time since each round of compensation will, all things being equal, reduce the residual unfair climate burden.

[15] Although we talk of 'policymakers' as the 'perpetrators' of secondary injustice and thus the bearer of the duty of compensation for secondary injustice, P could also be thought of as a state (or a group of states acting collaboratively to combat the injustice caused by climate change). It is also worth noting that, although responsible for injustice, the 'perpetrators' in our model are not necessarily wrongdoers (in the sense of being culpable) but rather creators of loss and damage that it would be unjust not to rectify. We therefore depart from the standard use of the term.

two factors are (1) *other alternatives* and (2) *inexcusable ignorance*.[16] Where either (and especially both) of these factors are present, a policymaker, P, can be said not to be acting fully innocently with respect to a victim of secondary injustice, V, such that V may reasonably demand additional compensation from P to match the increased moral damage done to them. We do not attempt to specify the exact magnitude of the compensation enhancements in these contexts but rather we simply introduce the two factors and explain why they should affect the overall amount of compensation due. To this end, we endorse a pluralist approach, accepting that the relative weighting of each factor may vary on a case-by-case basis.

3.2. Other alternatives

In some contexts, a policymaker, P, may only have a single course of action available to them that would discharge their duty of combating climate change. This will likely be the case if P is located in a state that is impoverished or especially vulnerable to the impacts of climate change. P, in such circumstances, may have little or no option other than to implement a policy that generates secondary injustice. Under the model we are proposing, P, due to their lack of options, would be excused from providing additional compensation to V above the 'fair burdens baseline' on the grounds that no one can be singled out for discriminatory treatment if the policymakers could not have done otherwise. All other things being equal, to require P to make V fully whole again (or to compensate above the fair burden baseline) would involve P being treated in the same way as we would treat a policymaker that could do more (but refused) to reduce secondary injustice. This, we argue, would be unfair on P and overstate the moral damage done to V. However, some compensation from P to V is due in recognition that P's action has damaged the life of V in a way that it has not damaged others and there is no justification for P to ignore V's continuing disadvantage in this respect. P might have done the right thing all things considered, but there is some moral cost which ought to be acknowledged through compensation indexed to the fair burden baseline.

In other cases, a policymaker may have a greater range of policies available to them. Where more than one option is available, P, has a duty to choose the *most just* option available. What counts as 'most just' will depend upon a broader account of justice. However, all of the options will involve an ethical balancing of the claims of all potential victims of secondary and primary climatic injustice. Imagine that there

[16] It might be objected that these two factors also act as constraints on the permissibility of the associated policy (had there been less disruptive or risky alternatives, and had policymakers known about these, then the policy would have been impermissible and so the injustice, on our own view, would be primary nor secondary) but there is nothing unusual in factors like these being used as compensation modifiers where the all things considered permissibility of the policy is not in question.

are four options available to P, with 1 being the most just and 4 being the least just. If P chooses 1, then it can be said that P has "acted in good faith" since P has done the best that could be expected of any policymaker similarly situated. P need not compensate V above the fair burden baseline to remedy the secondary injustice imposed on V in such a situation. However, if P chooses the second most just option in the available set, P has made V worse off than V relative to the most just policy option. This means that P has breached a duty of care owed to the victim and, even if the qualities of the policy are such that it remains a permissible choice all things considered, P now owes additional compensation in light of the additional damage done to V. Although we do not specify the exact boost to the compensation owed, it is natural to conceive it as bringing V up to the point where their condition is what it would be in a world where option 1 was chosen and its secondary costs were distributed equitably.

3.3. (In)excusable ignorance

As we have seen, due to the complexity of the climate change problem, policies that, on the available evidence, seemed to be the most just at the time of adoption may later be found to cause secondary injustice. The policymakers in such cases failed to select the best option for reasons beyond their control and not because they disregarded the interests of potential victims. This, we argue, may affect the amount of compensation owed from P to V. Imagine that P enacts option 1 believing it reasonably to be the best option, but ten years later, option 1 is found to have adverse effects that create far more secondary injustice than option 2. Option 2, not option 1, has become the most just option with the passage of time and yet P could not have reasonably known this at the time option 1 was selected. To require P to provide full compensation to V would be to treat P the same as a policymaker who culpably caused the injustice to V and yet P is not culpable since the injustice they caused was created not by malice or recklessness but by an "honest mistake." Full compensation would also treat V as if their disadvantage had arisen from wrongdoing which it did not. Instead, as argued above, V should be compensated so that they bear an equitable share of the secondary effects of policy 1. However, if P had succumbed to one form of moral corruption and ignored the available evidence, then P would have been culpable for V's worse position under the chosen policy; and they would also have inflicted a greater moral loss on V than had P acted innocently. Our claim is that P, in such circumstances, would owe V compensation beyond the 'fair burden baseline' so that V's condition is as it would have been, had the retrospectively most just option been selected (in this case, option 2), and the secondary effects of this option are shared by all.

3.4. "Honest mistakes"

The extent to which policymakers acted in "good faith" is key to our claim that compensation for secondary injustice may vary from case-to-case. It is therefore natural to ask whether such judgements can be reliably made. Obviously, we can never truly know what is in a policymaker's mind when they act. As such, any judgement of "good faith" would appear to be subjective, contestable, and impossible to implement fairly or consistently. Though it is clear that judgements of good faith, like any human judgement, are fallible and subject to contestation, it is important not to overstate the implication of this for our inquiry. In practice, agents –from individual human beings to institutions such as courts and scrutiny committees— make judgements of good faith on a daily basis. Like the term "reasonableness", we may never be able to set out necessary and sufficient conditions for such judgements being warranted but this does not prevent "good faith" from being deployed in moral theory and practice. In everyday moral practice, we look for an indication that agents considered available evidence and engaged in some serious conscientious reflection before acting (if they had time available). Demonstration of readiness to admit mistakes, and to offer redress for them, is also an indicator of good faith; doing so shows that an agent takes seriously the impacts of their actions and is more concerned about those than for example their own reputation and status.

In the case of institutional agents, we can go further and sketch what might count towards evidence of good faith in climate policymaking. First, there have already been suggestions made of conditions that contribute to "maladaptation" (Barnett and O'Neill 2010). Some of those, particularly path dependency, are also identified by social scientists who work on the development of new technologies, particularly carbon dioxide removal (CDR) and SRM methods.[17] A sincere attempt to take heed of this advice and avoid implementing policies with these conditions could be taken as an indicator of good faith. At the very least, it would be disingenuous for an agent to claim that they expected a project which exhibited or encountered these conditions to be problem-free.

Second, we could expect the conscientious reflection undertaken by policymakers to include consideration of the social scientific research about how individuals, institutions and social groups manage and knowledge and ignorance. For example, Steve Rayner introduces the term "uncomfortable knowledge" to refer to 'information or understanding that is available to certain parties, but cannot be acknowledged by others' (2012:113). Organisations, according to Rayner, typically cope with uncomfortable knowledge by one of four methods: denial, dismissal,

[17] See, for example, Cairns (2014). For suggestions on how to avoid maladaptation also see Mangan (2014).

diversion and displacement (2012: 113-122). However, the better way to manage uncomfortable knowledge is to ensure diversity of perspectives in the decision-making process (2012: 123). We suggest that a concerted effort to gather all available information about courses of action, including those emanating from marginalised perspectives, can be regarded as an indicator of good faith.[18] Being prepared to listen to others' perspectives and to be flexible shows that an agent is trying to find a workable solution rather than driving through initiatives regardless of the consequences. In summary, it is possible to make meaningful, albeit fallible, judgements about an agent's good faith and, in the case of institutional agents, taking seriously the idea of "wicked problems" is a useful indicator of good faith.

3.5. Constrained compensation and "rough justice"

We have argued that all things being equal, a policymaker (P)'s duty to provide compensation to a victim of secondary injustice, V, extends beyond the 'fair burdens baseline' if (a) P had better policies available to them than the one adopted and/or (b) P failed to act in good faith in the sense that they failed to choose what they reasonably believed to be a better option. These compensation enhancing conditions may seem to be discrete but, in reality, both are matters of degree and hence the boost (or not) they make to V's claim of compensation against P will also be a matter of degree. Where the two conditions instantiate perfectly, the policy in question will almost certainly be impermissible and the policymakers will owe full compensation to victims of secondary injustice regardless of any successes in managing climate change. Where the two conditions do not arise at all, the duty of policymakers to compensate victims of secondary injustice is sharply constrained in line with P's lack of culpability for V's impaired state. Rather than give *full* compensation to V (to restore V's condition to the level he or she was before the policy was implemented), P would have to provide *some* compensation and this is naturally seen as benefiting V so that they experience a fair share of the additional costs of the policy rather than suffering disproportionately.

Finally, situations where the two conditions are present to some degree are perhaps the most interesting and most commonly arising. Here, compensation is owed somewhere between the extremes of full compensation and the minimal compensation secured by the fair burdens baseline. But this should be seen not as a troubling result but rather an intuitive advantage of recognising the uniqueness of injustices arising from just policies. Because secondary injustices caused by agents acting in imperfectly good faith may still be, and often will be, *just* interventions, there is no

[18] This is also a key feature in the design of "clumsy" or "loose fit" solutions, which are thought to be more effective in solving "wicked problems" such as climate change (Verweij and Thompson 2006).

requirement to treat the associated victims as if they were victims of primary injustice demanding nothing less than full compensation. This way of thinking about secondary injustice places much weight on our judgements about the "good faith" exercised by policymakers when they select between the options available to them and allows for significant gradations in good faith. It is unlikely that a rigid model that sets out clear priorities and formulas for calculating what is owed to victims of secondary injustice is possible. Instead, the aim should be to bring about a sort of "rough justice" amongst the parties.[19]

4. Conclusion

The objective of this paper was to introduce the concept of secondary climatic injustices and to set out a case for considering their just rectification. We argued that, once policies are distinguished as permissible or impermissible, there are three different types of secondary injustice arising from otherwise permissible policies. Each of those three types can arise for any preventative response measures. Like primary climatic injustices, secondary injustices are likely to disproportionately burden the poor, the vulnerable and future generations and so the moral impetus for addressing them is the same. However, secondary climatic injustices differ from primary climatic injustices in that the latter arise in many cases from "good faith" attempts to lessen the injustice of climate change. Where such "good faith" is fully present, we argue that agents responsible for secondary climatic injustices may contribute substantially less than those responsible for primary climatic injustices but never less than what would bring the victims up to the point where they no longer bear an unfair share of the burdens of secondary injustice. It was not possible to present a full theory of rectification for secondary injustice and it may not be possible to develop a highly principled, technical theory, due to the complexity of the issue. We suggested that an appeal to constrained compensation, itself grounded in what others have called "rough justice," may have to suffice. However, we presented two criteria which we believe can render more exact the amount of rectification required.

[19] Linzer (2001: 695) defines rough justice 'as driven more by general standards of fairness than by structured (or formal) systems of rules and neat categories, justice that is often untidy, that may be second-best where the best is unachievable.'

References

Barnett, Jon and O'Neill, Saffron (2010), 'Maladaptation', *Global Environmental Change,* 20 (2), 211–13.

Barnett, Jon and O'Neill, Saffron J. (2013), 'Minimising the risk of maladaptation', *Climate Adaptation Futures* (John Wiley & Sons), 87–93.

Byskov, M.F., Hyams, K. *et al* (2021) 'An agenda for ethics and justice in adaptation to climate change', *Climate and Development*, 13 (1): 1–9.

Cairns, Rose C. (2014), 'Climate geoengineering: issues of path-dependence and socio-technical lock-in', *Wiley Interdisciplinary Reviews: Climate Change,* 5 (5), 64–61.

Caney, Simon (2005), 'Cosmopolitan Justice, Responsibility and Global Climate Change', *Leiden Journal of International Law,* 18 (4), 747–75.

Caney, Simon (2011a), 'Climate Change, Human Rights and Moral Thresholds' in Climate Ethics: Essential Readings (eds Gardiner, S., Caney, S. Jamieson, D., and Shue H. Oxford: Oxford Univerisyt Press), 163–177.

Caney, Simon (2011b), 'Just Emissions', *Philosophy and Public Affairs,* 40 (3), 255–300.

Chan, Nicholas (2016) 'The "New" Impacts of the Implementation of Climate Change Response Measures' *RECIEL* 25(2), 228–237.

Dow, Kirstin, et al. (2013), 'Limits to adaptation', *Nature Climate Change,* 3 (4), 305–07.

Eckersley, Robyn (2015), 'The common but differentiated responsibilities of states to assist and receive 'climate refugees'', *European Journal of Political Theory,* 14 (4), 481–500.

Eriksen, Siri, et al. (2011), 'When not every response to climate change is a good one: Identifying principles for sustainable adaptation', *Climate & Development,* 3 (1), 7–20.

Fischer, Wolfgang (2015), 'No CCS in Germany Despite the CCS Act?', in Wilhelm Kuckshinrichs and Jürgen-Friedrich Hake (eds.), *Carbon Capture, Storage and Use: Technical, Economic, Environmental and Societal Perspectives* (Cham: Springer International Publishing), 255–86.

Gardiner, Stephen M. (2010), 'Is "Arming the Future" with Geoengineering really the Lesser Evil? Some Doubts about Intentionally Manipulating the Climate

System', in Stephen M. Gardiner, et al. (eds.), *Climate Ethics: Essential Readings* (Oxford: Oxford University Press), 284–312.

Goodin, R. E. (1985), *Protecting the Vulnerable: A Reanalysis of Our Social Responsibilities* (Chicago: University of Chicago Press).

Goodin, Robert E (1991), 'Compensation and Redistribution', in J.W. Chapman (ed.), *Nomos XXXIII Compensatory Justice* (New York: New York University Press), 143–77.

Heyward, Clare (2013), 'Situating and Abandoning Geoengineering; A Typology of Five Responses to Climate Change', *PS: Political Science and Politics,* 46 (1), 23–27.

Hulme, Mike (2009), *Why We Disagree about Climate Change: Understanding Controversy, Inaction and Opportunity* (Cambridge: Cambridge University Press).

Hunold, Christian and Young, Iris Marion (1998), 'Justice, Democracy, and Hazardous Siting', *Political Studies,* 46 (1), 82–95.

Huq, Saleemul, Roberts, Erin, and Fenton, Adrian (2013), 'Loss and damage', *Nature Climate Change,* 3 (11), 947–49.

IPCC (2014) 'Summary for Policymakers', in C.B.Field (ed) *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Working Group II Contribution to the IPCC 5th Assessment Report*. Geneva: United Nations, pp.1-44. Available at: http://ipcc-wg2.gov/AR5/images/uploads/IPCC_WG2AR5_SPM_Approved.pdf.

IPCC (2023) 'Summary for Policymakers', in H. Lee and J. Romero (eds) Climate Change 2023: Synthesis Report. Geneva: United Nations, pp. 1-42. Available at: https://www.ipcc.ch/report/ar6/syr/downloads/report/IPCC_AR6_SYR_SPM.pdf

Linzer, Peter (2001), 'Rough Justice: A Theory of Restitution and Reliance, Contracts and Torts, *Wisconsin Law Review* 696–775.

Miller, David (2007), *National Responsibility and Global Justice* (Oxford: Oxford University Press).

Moellendorf, Darrel (2014), *The Moral Challenge of Dangerous Climate Change* (Oxford: Oxford University Press).

Morrow, David R (2014a), 'Ethical aspects of the mitigation obstruction argument against climate engineering research', *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences,* 372 (2031), 20140062.

Morrow, David R. (2014b), 'Starting a Flood to Stop a Fire? Some Moral Constraints on Solar Radiation Management', *Ethics, Policy & Environment,* 17 (2), 123–38.

Noble, I. R., et al. (2014), 'Adaptation needs and options', in C. B. Field, et al. (eds.), *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel of Climate Change* (Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press), 833–68.

Nuffield Council on Bioethics (2011) *Biofuels: Ethical Issue*s. London:Nuffield Council on Bioethics.

O'Neill, Brian C. and Oppenheimer (2004), 'Climate change impacts are sensitive to the concentration stabilization path', *Proceedings of the National Academy of Sciences,* 101 (47), 16411–16.

Page, Edward and Heyward, Clare (2017) 'Compensating for climate change loss and damage', *Political Studies* 65(2): 356–72.

Rayner, Steve (2012), 'Uncomfortable knowledge: the social construction of ignorance in science and environmental policy discourses', *Economy and Society,* 41 (1), 107–25.

Rayner, Steve, et al. (2013), 'The Oxford Principles', *Climatic Change,* 121 (3), 499–512.

Ripstein, Arthur. 2007. 'As If It Never Happened.' *William and Mary Law Review*, 48: 1957–97.

Shepherd, John, et al. (2009), 'Geoengineering the Climate: Science, Governance and Uncertainty.', (London: The Royal Society).

Sidi, Purnomo (2012), 'Integrated Climate Action: Linking Mitigation and Adaptation to Make Indonesian Cities Resilient', in Konrad Otto-Zimmermann (ed.), *Resilient Cities 2: Cities and Adaptation to Climate Change – Proceedings of the Global Forum 2011* (Dordrecht: Springer Netherlands), 349–57.

UNFCCC (2012) 'A literature review on the topics in the context of thematic area 2 of the work programme on loss and damage: a range of approaches to address loss and damage associated with the adverse effects of climate change.' Geneva: UNFCCC. Available at: http://unfccc.int/resource/docs/2012/sbi/eng/inf14.pdf.

Vanderheiden, Steve (2008), *Atmospheric Justice: A Political Theory of Climate Change* (Oxford: Oxford University Press).

Vermule, Adrian (2012), 'Reparations as "Rough Justice"', in Melissa S. Williams, Jon Elster, and Rosemary Nagy (eds.), *Transitional Justice* (New York, NYU Press)

Verweij, Marco, et al. (2006) Clumsy Solutions for a Complex World: The Case of Climate Change', *Public Administration,* 84 (4), 817–43.

Wallimann-Helmer, Ivo (2015), 'Justice for climate loss and damage', *Climatic Change,* 133 (3), 469–80.

Robert Huseby[1]

Sufficiency and the Distribution of Burdens[2]

A common objection to sufficientarianism is that it allows large inequalities above the threshold. A sharpened form of this objection highlights that this indifference also encompasses large inequalities in the distribution of burdens. Consider the burdens that follow from climate change. A theory that does not rule out placing these burdens on the worst off (of the sufficiently well off) will appear implausible to many. This paper assesses ways of addressing this objection and defends a revised conception of sufficientarianism that can demand fair distribution of burdens (and benefits) above the sufficiency threshold, without giving up core sufficientarian theoretical commitments.

[1] Robert Huseby,Department of Political Science, University of Oslo (roberthu@uio.no)

[2] Acknowledgements: This paper was presented at the Nordic Network for Political theory in Tromsø in 2022, and at the PPE seminar at the Institute for Futures Studies in Stockholm in 2022. I am grateful to Kim Angell, Krister Bykvist, Naima Chahboun, Jakob Elster, Göran Duus-Otterström, Hilary Greaves, Fredrik Dybfest Hjorthen, Karim Jebari, Julia Mosquera, Frank Nullmeier, Daniel Ramöller, Thomas Schramme, Jørn Sønderholm, Jens Damgaard Thaysen, Olle Torpman, and two anonymous reviewers for very helpful comments

1. Introduction

Distributive justice typically concerns the just distribution of various kinds of *goods*.[3] If we know which goods are relevant for justice, we can divide them up according to the most plausible distributive principle. If you are an egalitarian, you will distribute the relevant goods equally. If you are a sufficientarian, you will distribute them so as to maximize the number of people that reach the sufficiency threshold, or to minimize the amount of insufficiency (that is the (morally weighted) aggregate shortfall from the threshold).[4]

Distributive justice also, arguably, concerns the distribution of *burdens*. Burdens can come in many forms but I will mainly be concerned with the kind of burdens that are in some sense unavoidable, regardless of whether or not they are ultimately caused or imposed by human agency. More particularly, consider burdens that are like the burdens that follow from climate change. Climate change already causes harm to many and will continuously and increasingly do so in the foreseeable future. Temperatures will increase, sea levels will rise, deserts will spread, contagious diseases will become more common, millions will be forced to flee their homes, and so on.

These burdens will most likely fall unevenly and unfairly on people if nothing is done to secure a fair distribution (Caney 2010). Note that the term *burdens* is here used in a special sense. I assume that burdens are to a large extent fungible and sharable. To be sure, for a small island state disappearing into the rising sea, it may not make sense to think of that (gigantic) loss as fungible and sharable. Nonetheless, there are things that others can do to share this burden at least to some extent. The islanders can be provided with a new territory, they can be compensated financially, granted entry to other states, and so on. I do not claim that the loss of a territory can be fully compensated, or that the burdens associated with such a loss can be shared in full, but I take it that it can be shared to a large degree.[5]

More generally, I assume that most burdens following from climate change can be shared, in some way or the other, either at the mitigation stage, at the adaptation stage, or at the compensation stage (if and when adaptation in some sense is no

[3] In this paper, I will use 'goods' as a placeholder for whatever it is that we think should be distributed justly.

[4] Views that imply maximizing sufficiency (Frankfurt 1989) have the troubling implication that if everyone is very badly off, it is better to have one person reach the threshold, than to have everyone *almost* reach it (Casal 2008). I will for that reason set such maximizing versions aside for the rest of the paper.

[5] Caney 2014 distinguishes between harm avoidance and burden sharing, because sometimes harms must be avoided immediately, before a scheme of burden sharing is in place. I am here, as suggested above, concerned with how to share those burdens fairly. I thus leave the ethics of harm avoidance to one side.

longer possible, such as in cases of lost territory). This is in line with most writings on climate justice (Caney 2005, 2010, 2014, Page 2008, Jagers and Duus-Otterström 2008, Gosseries 2003).

Further, it seems clear that these burdens should be distributed fairly. There are many discussions of burden sharing in relation to climate change. Since climate change is largely caused by human agency, several of the principles that are discussed are principles of *corrective* justice, that is principles that are invoked when some moral agent has inflicted harm, violated a duty, or in other ways incurred particular obligations to contribute to the rectification of a wrong. In the literature, the polluter pays principle (PPP) and the beneficiary pays principle (BPP) have both been extensively debated.

The PPP, as the name suggests, holds that those who are responsible for polluting should pay the cost, or take on the burdens, that follow from their emissions. This principle, then, can be seen as a corollary to a form of harm principle, according to which one should compensate for the harm one has inflicted, in this case through pollution (Caney 2005, Page 2008).

The BPP, on the other hand, has it that those who have benefited, even innocently, from injustice (or sometimes harm), have special duties to compensate those that have been harmed by the injustice, in this case pollution (Caney 2005, Page 2008, Haydar and Øverland 2015). However, for various reasons (dead polluters, innocent pollution, the non-identity problem, natural climate change) principles of corrective justice, justifiable or not as such, do not apply to (anything near) the total burden that follows from climate change (Caney 2005; 2010). A substantial portion of this burden, then, must therefore be distributed in a more general and perhaps consequentialist fashion, similarly to burdens that are not caused by human agency, for instance (pure) natural disasters. The ability to pay principle (APP) is one such suggestion. According to this principle the (remaining) burdens that follow from climate change should be distributed in light of the relevant agents' capability to shoulder them (Caney 2010).

Other principles that are discussed in the broader literature on distributive justice, are more general, and not limited to corrective justice (PPP and BPP) or the burdens that remain after principles of corrective justice have been applied (APP). These more general theories, such as egalitarianism, prioritarianism, utilitarianism, sufficientarianism, and limitarianism may also be applied to climate change. Some (versions) of these have little or no place for corrective principles such as PPP and BPP (for instance total utilitarianism and total prioritarianism). Others, such as for instance forms of egalitarianism and sufficientarianism, are compatible with them, but will be competitors to APP (which, as noted, applies to the portion of the burden that cannot be assigned in light of more standard corrective principles).

In my view, *sufficientarianism* is a promising theory of distributive justice in many respects. Sufficientarianism, in most renditions, defines a threshold that signifies a level of goods that it is morally important that people reach, and rejects further distributive principles above that threshold. However, sufficientarianism faces a quite serious objection - the indifference objection - according to which this theory allows large, potentially extremely large, inequalities above the threshold (Casal 2008, Knight 2022).[6] A sharpened form of the indifference objection points to the fact that sufficientarianism is also unable to distribute *burdens* (those that are not distributable in light of PPP or BPP, to the extent that the version of sufficientarianism in question is compatible with these principles) in a fair manner. This objection is problematic not least in the context of climate change, where burden sharing comes into particularly sharp focus. Many agree that a theory that is unable to distribute the burdens of climate change fairly, is implausible. Thus, it would significantly strengthen sufficientarianism if the burden objection could be met.

This paper assesses different ways of addressing the burden objection (some of which, by extension, also addresses the indifference objection in its entirety), and defends a revised conception of sufficientarianism that requires fair distribution of burdens (and benefits) above the sufficiency threshold, but without giving up core sufficientarian theoretical commitments. Thus, the aim is to strengthen sufficientarianism, and in that way support the theory. The aim is not to undermine or criticize other theories, or to assess the extent to which sufficientarianism is overall more plausible than its rivals.

2. Sufficientarianism and the Burden Objection

Sufficientarianism comes in many different forms (Axelsen and Nielsen 2015, Benbaji 2005, Crisp 2003, Frankfurt 1987, Huseby 2010, 2020, Shields 2011, 2016, Timmer 2022),[7] but I will for the time being just assume a simple version, according to which there is a threshold of goods such that it is especially important that people reach it. Above this threshold, no further principles of distributive justice apply. These two claims correspond to sufficientarianism's positive thesis (PT) and negative thesis (NT), respectively (Casal 2007).

[6] This objection is sometimes presented in terms of the distribution of benefits, in particular (Casal 2007), and sometimes in terms of distribution above the threshold in general (Knight 2022). The latter version encompasses the former version as well as the burden objection (see below).

[7] Sufficientarianism has also been subject to a range of criticisms and objections that will not be addressed here. For further discussion, see Arneson 2005, Casal 2008, Herlitz 2018, Knight 2022, Nielsen 2016, Segall 2016, Widerquist 2010.

PT: There is a level of goods such that it is especially morally important that people reach it.

NT: There is a level of goods such that above it, no distribution is unjust.

The negative thesis might cause some trouble for sufficientarians. Suppose everyone is above the threshold, but that there is inequality between two groups. One group is just above the threshold, say, and another group is substantially above it. For simple illustration, suppose the threshold is at 10, that the worst-off group is at 11, the best-off group is at 21, and that there are equally many individuals in each group.

In light of NT, this situation does not give rise to any claims of redistribution in the name of justice. So long as everyone is sufficiently well off, everything is fine, distribution-wise. As indicated, many critics find this objectionable in itself (hence the indifference objection). Worse still, however, suppose that some substantial burden presents itself. This burden is similar to that portion of the burden of climate change that cannot be distributed in light of any plausible corrective principles of justice (such as PPP or BPP). Thus, there is no morally salient connection between the agents and the burden that give any of them particular duties to shoulder it. None of them have caused the burden, and none of them will benefit from its landing on others.[8] We can further assume that, for some reason, we can only distribute this burden in one of two ways. Either we can place it on the worst off, or we can place it on the best off. For illustration, consider a simplified world in which the remainder (the part of the climate change burden that cannot be distributed in light of corrective principles of justice) will land on the worst off if nothing is done. Even if they are the worst off, however, we assume that they are still sufficiently well off by a relatively slim margin. A sustained effort on the part of the best off will protect the worst off from this burden, but this effort will be relatively costly. The best off are, however, (financially) more than capable of taking it on. In both cases, the cost equals one unit of goods per person, such that either the worst off end up with 10, that is, exactly at the threshold, or the best off end up with 20, that is, less than they used to have, but still quite a lot and still well above the threshold.

In this case, it seems that it would be wrong to place the burden on the worst off, even if they are sufficiently well off. Presumably, many sufficientarians would

[8] I make this specification because sufficientarianism, as mentioned, may well be compatible with principles of corrective justice, such as the BPP. Note that even if a general distributive theory may be compatible with principles of corrective justice, conflicts between them can still arise in specific circumstances. Thus, some sort of priority or weighting will be required.

agree.[9] To be sure, sufficientarianism does not imply that we *ought to* burden the worst off, but the theory is in most versions *indifferent* between burdening the worst off and burdening the best off (so long as all remain sufficiently well off). Worse still, standard sufficientarianism is indifferent between burdening the worst off, and not burdening *anyone*. If, for some reason, we can avoid an impending catastrophe either by substantially burdening the worst off, or by costlessly pushing a button, it is hard to see what reasons we have *qua* sufficientarians, to just push the button.[10]

This apparent inability of sufficientarianism to rule out forms of intuitively unjust (sometimes *very* unjust) ways of sharing burdens, including, it has been claimed, regressive taxation, is a very forceful objection to the theory (Kanschick 2015, Nielsen 2019). Burdening the worst off simply goes against the grain of what most people associate with the very concept of justice. Letting the worst off take the cost for avoiding harm to all, would in a sense be similar to taking from the poor (relatively speaking) and giving to the rich.[11]

Nielsen (2019) contrasts two cases that bring out the normative difference between what he refers to as the benefit-driven and the burden-driven versions of the indifference objection. In the *Manna from heaven* case, the imaginary island *Plentia* comprises a population in which everyone is sufficiently well off, but one third - the rich - are much better off than the rest. Then, some act of nature occurs that makes the rich even better off than they were, without affecting the remaining two thirds (Nielsen 2019: 27). In the *Manna from Hell* case, on the other hand, the act of nature makes the non-rich worse off than they were, even if they remain sufficiently well off, while leaving the rich unaffected (Nielsen 2019: 31).

According to Nielsen, the latter case appears more unjust than the former, even though the resulting inequality is the same (we can assume), and even if the worst off remain sufficiently well off. One reason might be that it is easier to imagine the

[9] Some sufficientarians might not find it problematic to place an unavoidable burden on the worst off rather than the best off, when everyone is sufficiently well off. These sufficientarians will, faced with the burden objection (and the indifference objection) simply bite the bullet. I assume that this is the minority position.

[10] I am grateful to Göran Duus-Otterström for suggesting this point.

[11] Note that PT might also have some potentially counter-intuitive implications. Suppose prioritarianism is applied below it (as is the case for several sufficientarian theories). If so, a burden might have to be placed on the worst off in cases where the total (morally weighted) disutility of doing so is less than the total (morally weighted) disutility of placing it on the best off (of those who are under the threshold). This is similar to the implications of ordinary prioritarianism. The difference is that ordinary prioritarianism will balance benefits and burdens among individuals *regardless* of whether they are above or below some sufficientarian threshold. I will not address this potential problem for PT here, but I do think that this is a smaller problem than the one raised by NT, because in the latter case, as noted above, the worst off can in principle be burdened even if the total disutility of doing so is *less* than burdening the better off. In addition, applying prioritarianism below PT, while problematic in some cases, is the most plausible alternative to applying leximin, which would raise other severe problems.

non-rich two-thirds being unmoved by others good luck, than with their own bad luck (Nielsen 2019: 31). I agree, and the contrast comes out even starker in cases where the worst (but sufficiently well) off must shoulder a burden in order to benefit all. Suppose a flood is threatening the whole of *Plentia*, and that expensive flood walls must be built. It would not appear unjust to let the rich one third take this cost. Sharing the cost among all might also be acceptable. But placing the whole cost on the worst off seems manifestly unjust.

I agree then, with Kanschick (2015), Nielsen (2029), and others, that the burden objection is even more forceful than the indifference objection, and moreover, that it is likely to be the most challenging objection to sufficientarianism altogether. As noted, it seems to me that the burdens following from climate change underscores this view. It is more pressing to avoid that these burdens harm the worst off, even when they are sufficiently well off, than to generally reduce inequalities above the threshold.

Some attempts have been made at addressing this issue (Huseby 2010, Kanschick 2015), but it is unclear how successful these attempts are, and the question clearly merits further discussion. It should be noted, however, that my proposed answer to the burden objection (unlike some of the alternatives I discuss along the way) does not at all *depend* on the assumption that it is worse than the indifference objection.

The burden objection (and the indifference objection more generally) presents sufficientarianism with a dilemma of sorts. Either sufficientarianism cannot meet the objection, and is, for that reason, implausible, or sufficientarianism can meet the objection, but only at the cost of rejecting NT, and thus losing its distinctiveness (to an extent at least). The idea behind the second horn, is that meeting the objection requires the acceptance of some further non-sufficientarian distributive principle above the threshold, something that NT seems to explicitly deny. And to the extent that one agrees with Casal (2003) that sufficientarianism is distinctive partly in virtue of accepting NT, the second horn follows.

To be sure, the second horn is only problematic to the extent that sufficientarians think NT is valuable, but it seems that many do. There are various reasons why that may be the case, but in my view (which I will elaborate below), NT is valuable because it (on some versions at least) points to a threshold such that below it, people have *absolute* priority over those above it. This blocks certain forms of aggregation, including aggregation that allows many small benefits to the very well off to outweigh a few large burdens to the very badly off. (This kind of aggregation, moreover, is what many see as the main problem with theories such as utilitarianism and prioritarianism.)

There are at least three ways of addressing this issue. The first is to deny the force

of the burden objection. I will not pursue that alternative here. It seems, as I have suggested throughout, that the objection is very forceful indeed, and that sufficientarianism will be strengthened to the extent that it is able to meet it, rather than just bite the (distasteful) bullet.

Another way out of the dilemma is to argue (as some have argued), that there are *instrumental* reasons to distribute burdens fairly or progressively, even above the threshold. Instrumental reasons in this context, are reasons that do not flow from intrinsically valuable distributive principles, such as equality or priority (or sufficientarianism), for instance. If, say, inequalities above the threshold tended to threaten the long-term sufficiency of the worse off, or if inequalities above the threshold tended to undermine social stability (and hence sufficiency) we would have such instrumental reasons not to place burdens on the worst off (and also not to benefit the better off). Redistribution from the better off to the worse off above the threshold would not be intrinsically good (as many egalitarians would claim). Rather, equalizing would be instrumentally good in virtue of securing sufficiency over the long term, which (on this account) would be intrinsically good. I will consider some attempts along these lines in section 3 below.

A third possibility is to *redefine* NT in such a way that it allows for further distributive principles. I will consider and defend such an alternative (in section 4 below). More specifically, as suggested above, I will propose that we can restate NT in such a way that it allows further distributive principles above the threshold, but without thereby undermining the distinctiveness of sufficientarianism.[12] This restatement is based on the observation that NT really serves two purposes in most sufficientarian theories. First, it marks out the threshold above which no further distributive principles apply. Second it marks out the threshold at which those below it have *absolute* priority over those above (Huseby 2020). The latter purpose, importantly, is not in conflict with additional principles above the threshold, and it is sufficient to retain the distinctiveness of sufficientarianism, or so I argue. Moreover, the latter purpose is not dependent on the first, so that sufficientarians may well accept absolute priority below the threshold while accepting further distributive principles above. This solution, if successful, could meet both the burden objection, and the indifference objection more generally, because allowing additional distributive principles above the threshold would work equally well for benefits as for burdens.[13]

[12] Shields' shift-sufficientarianism (2016) can be seen as an attempt along similar lines. According to Shields, there is a threshold such that it is especially important that people reach it, but above that threshold, it might still be valuable to benefit people, but at a slower rate. The threshold thus marks a 'shift,' and Shields consequently rejects NT rather than redefine it. Crisp's version of sufficientarianism (2003) is probably compatible with my proposal. However, since he applies (or at least suggests) utilitarianism above the threshold, his view is unlikely to rule out all instances of *unfairness*.

[13] An alternative could be to show that there is a morally relevant *asymmetry* between the distribution of

3. Instrumental Reasons for Fair Burden Sharing

As noted, some attempts have been made to meet the burden objection by referring to instrumental reasons. Consider again the case above, where a burden could be distributed so as to either reduce the level of half the population from 21 to 20, or from 11 to 10 for the other half, and where 10 equals the sufficiency threshold. Sufficientarians could reasonably argue that laying the burden upon the worst off is objectionable because it makes half the population (even more) *vulnerable to insufficiency*, because they are after they have paid the cost of the burden just *barely* sufficiently well off (Kanschick 2015).

One problem with this, as an answer to the burden objection, is that it is less persuasive in cases in which half are at, say, 19 and the other half at 30.[14] Placing the burden on the worst off and reducing their level of wellbeing to 18 hardly puts them at risk of insufficiency. Of course, in most realistic scenarios, being at 18 makes one *more* vulnerable to insufficiency than being at 29, but not necessarily to any significant extent. In addition, it will, in principle, always be a contingent matter how vulnerable a person is to become insufficiently well off. Even if there is, empirically, a strong correlation between one's level of goods and one's vulnerability to insufficiency, the relationship will simply not hold in all cases (something that Kanschick explicitly acknowledges).

Now, consider the argument that we have instrumental reasons to avoid *large inequalities* (Scanlon 2018). If so, we would have reasons to avoid placing burdens on the worst off (of the sufficiently well off). There could be several possible reasons why inequalities are instrumentally bad, also from a sufficientarian point of view. For instance, inequalities could over time undermine social solidarity, or increase conflict.[15]

This argument, to the extent that it is empirically correct, seems perfectly fine, but it is not sufficiently general. If one group is at 19 and the other at 20, the inequality will not necessarily become large or problematic just by reducing the

---

burdens and goods. After all, the burden objection seems to presuppose something like that. If asymmetry turns out to be plausible, it appears that sufficientarians would be free to distribute burdens in a principled way, above the threshold, in light of some non-sufficientarian principle (for instance equality or priority), without thereby denying (the original) NT completely. That is, (the original) NT could still hold for the distribution of *benefits*. I leave this alternative to one side here, however, because it seems all in all better to be able to meet both the burden objection and the indifference objection.

[14] To be sure, when the worst off are at such a high level, the intuitions might change, and the burden objection might appear less problematic. As noted, however, I assume that most would find the burden objection problematic (which is compatible with thinking that it is more problematic the lower the level of the worst off).

[15] Instrumental reasons might come in many different versions, and some may be stronger than others. Here, I just assume, for the sake of argument, that there are some weighty instrumental (non-sufficientarian) reasons that can be invoked in order to avoid placing burdens on the worst off.

level of the worst off by one, to 18. (To be sure, this will depend on the facts that obtain in the different contexts). Also, if a third of the population is at 20, a third at 25, and a third at 30, it is not clear that the instrumental concern for equality is sufficient to rule out placing the burdens on the middle group.

Another possibility is to refer to relative deprivation (Huseby 2010). If the sufficiency threshold is specified (at least in part) with reference to subjective contentment, it could be the case that inequalities above the threshold are problematic because they (actually) lead to insufficiency, via relative deprivation. Suppose the current sufficiency level is at 10. A is at 15, and all others are at 20. If a burden is then placed on A, such that she ends up at 10, it could be the case that 10 is no longer sufficient, because A will (reasonably) evaluate her level of goods at least partly, in light of what levels of goods others have. (Even if, say, 10 would have been sufficient if everyone else was also at 10, or only A was at 10, but all others were at 12). This argument too, however, lacks generality. Whether or not people experience relative deprivation is essentially a contingent question.

I do not want to deny that instrumental arguments of this kind can be plausible, and successful in avoiding many actual forms of seemingly unfair burden-sharing, However, these arguments are, as indicated, contingent, and they are not general enough to rule out all cases of apparently unfair burden sharing. It would clearly be better to have a more general and principled argument in favor of fair burden sharing above the threshold.

It is also worth emphasizing that the solutions based on instrumental reasons have another flaw, in addition to their contingency and lack of generality. They all fail to meet the burden objection in a satisfactory way. The objection holds that some forms of distribution of burdens above the sufficiency threshold is *unfair*. The answers discussed above attempt to show that such unfair distributions will seldom be called for. But they fall short of deeming them unfair as such. Thus, even if they were more general and less contingent, they would still fail to answer the objection in a completely adequate manner.

4. Redefining the Negative Thesis

In my view, the best way for sufficientarians to handle the burden problem, is to redefine NT and accept a certain form of pluralism. This solution is to some extent foreshadowed in the literature, but has so far not been explicitly used to address the fairness of burden sharing (see Crisp 2003, Casal 2007: 300).

Consider NT as it has ordinarily been presented in the literature. According to Casal, the negative thesis 'denies the relevance of certain additional distributive requirements' above the sufficiency threshold (2007:298). Later on, she suggests

that it is in particular equality and priority that sufficientarians reject (2007: 299), but that sufficientarians tend to, and have reason to, reject other principles as well. For my purposes, this understanding of the negative thesis can be completely general and encompass all further distributive requirements.[16] Thus defined, the negative thesis conforms to the specification given above:

> NT: There is a level of goods such that above it, no distribution is unjust.[17]

On this definition, neither benefits nor burdens can be distributed in light of some (intrinsic) principle of fairness above the threshold, since no distribution above the threshold can be unjust. Unless a sufficiently general instrumental argument can be given for why we should prioritize the better off over the worse off, the choice between progressive and regressive taxation, for instance, remains a mere toss-up.

Thus, we should consider alternatives. One option would be to emphasize not the rejection of other principles, but, as I have already indicated, the *priority* of those below the threshold, over those above (Huseby 2020:211). The point is that on many sufficiency views, the negative threshold has two distinct functions. The first is the one associated with NT above, that is to deny the relevance of further distributive principles. The second is to assert absolute priority to those below the threshold over those above.[18]

It is only the first of these functions that make sufficientarianism vulnerable to the burden objection. The second does not. Also, the second, while in my view important, is not strictly speaking implied by NT as stated. In light of this, I suggest the following, alternative, version of NT:

> NTa: There is a level of goods such that those below it have absolute priority over those above it.

There are some things to note here. First, one might think that the absolute priority to those below the threshold is a feature of *PT*, rather than NT. In Casal's (and for that matter, others') formulation, that is not the case. Recall,

[16] But note that this does not exclude catering to deontological concerns such as respect for individuals (see Frankfurt 1989, Casal 2007).

[17] For a different formulation, see Shields 2012:103.

[18] Note that Limitarianism also implies a version of the negative thesis. According to this theory (in at least some renditions) it is impermissible to have goods above a certain threshold (Robeyns 2017). Limitarianism and sufficientarianism are, however, distinct theories.

> PT: There is a level of goods such that it is especially important that people reach it.

PT can be specified in different ways, however, as pointed out by, among others, Casal (2007). For instance, it *can* imply absolute priority or (merely) strong priority to those below the threshold. Some sufficientarians have defended absolute priority (Crisp 2003, Frankfurt 1987, Huseby 2010), whereas others have defended non-absolute versions (Axelsen and Nielsen 2015:418, Shields 2011:107). Now, suppose one defends an absolute version of PT:

> PTa: There is a level of goods such that it is especially important that people reach it. People below it have absolute priority over people above it.

Thus defined, PTa seems to *entail* the alternative understanding of NT above.[19] So, for sufficientarians that are not terribly concerned with outright denying the possibility of further distributive concerns above the threshold, an absolutist understanding of PT would in fact be enough to define a recognizable and distinct sufficientarian position.

To be sure, some have claimed that sufficientarianism proper implies a commitment to both theses. But it seems that any position that holds an absolutist version of PT (PTa) would be distinctly sufficientarian (admittedly, partly in virtue of entailing NTa). After all, no other familiar distributive principle gives absolute priority to all who are below a sufficiency threshold. But it would of course be easier to *combine* such versions with further principles above the threshold, than forms of sufficientarianism that embrace both theses, at least to the extent that NT is understood in Casal's manner.

However, if NT is understood as primarily stressing absolute priority (that is, NTa), forms of sufficientarianism that accept both theses can accept further principles as well. One might wonder, however, why one would accept NTa in the first place, if one already accepts an absolutist version of PT. That could seem superfluous. As I will argue next, there could be reasons for accepting NTa, but it is worth emphasizing that both possibilities would have the resources to meet the burdens-objection.

If we suppose that the two theses (however defined) refer to the *same* threshold, then it would indeed be superfluous to assert both PTa and NTa. PTa would define

[19] Furthermore, PT could be expressed identically to NTa. PTa/NTa: There is a level of goods such that those below it have absolute priority over those above it. This would be less informative, but still comprehensible. As I note in the main text, however, we have reasons to prefer to keep the two theses distinct.

the threshold *and* assert the absolute priority of those below it over those above it. To accept, *in addition*, NTa, which just repeats part of what is already asserted by PTa would serve no interesting purpose. However, if we assume that the two theses refer to *different* thresholds, the picture changes (Huseby 2020). In that case there is a need for an additional and distinct NTa, referring to a higher threshold.[20] Such a view is compatible with either a version of PT or PTa.[21]

Such a two-threshold view is compatible both with NTa and the original NT, but only versions accepting NTa could successfully meet the burden objection. To clarify, this view says

a) there is a level of goods such that it is especially morally important that people reach it,
b) there is a (distinct and higher) level of goods such that those below it have absolute priority over those above it.[22]

Interestingly, a) and b) are compatible with

c) above the higher threshold, equality (or priority) applies.

This would be a hybrid view, to be sure, but it would be distinctly sufficientarian nonetheless, in virtue of accepting NTa. (As noted the same can be said for alternatives that accept PTa.)

It seems then, that both versions (PTa, and PT plus NTa) can successfully avoid the objection from the distribution of burdens. (I happen to think that the latter version (PT plus NTa) is more plausible for other reasons but will not argue for that conclusion here.)[23] Since both are compatible with c) they can easily be combined

[20] There are, however, conceivable alternatives. For instance, there could be several PTa's and/or several NTa's. It will lead too far to canvass all possibilities here. I am grateful to an anonymous reviewer for raising this point.

[21] In principle, I assume that NTa could refer to a lower threshold, but I will not consider that alternative here.

[22] Note that this priority is intended to cover cases in which some individuals are at or just above the threshold, and risk falling below it, if some policy is implemented. These individuals would be given absolute priority over individuals who do not risk falling below. Below the threshold, more standard prioritarian reasoning apply. I am grateful for an anonymous reviewer for raising this issue.

[23] Note that this solution avoids one possible problem with the asymmetry thesis. According to asymmetry, burdens and benefits can be distributed in light of different principles. This might for all I have said so far be plausible, but it is hard to avoid the thought that it is difficult to maintain a conceptual distinction between the two. And even if it is possible to maintain such a distinction, some bona fide benefits might appear to be such as to call for a fair distribution. In addition, many have raised the same objection with reference to benefits (Casal 2007). Versions based on asymmetry cannot address such objections. Views that accept PTa or PT plus NTa can.

with further distributive principle above the (higher) threshold. If we suppose, for the sake of argument, that prioritarianism is chosen, burdens falling on individuals above the (higher) threshold, will be distributed in a way that prioritizes the worst off.[24] In addition to meeting the burdens objection, this solution also meets the indifference objection in its entirety, since this further distributive principle can apply to benefits as well as burdens. On this revised view, then, burdens in general, as well as the particular burdens following from climate change, can be distributed fairly.

5. Conclusion

In this paper, I have considered the burden objection to sufficientarianism. This objection holds that sufficientarianism allows unfair burden sharing. This seems problematic. In response, I have outlined a form of sufficientarianism that is immune to this objection, in virtue of accepting a redefined version of the negative thesis, that emphasizes absolute priority to those below the threshold, rather than the rejection of further distributive principles above the threshold. In my view, such a version of sufficientarianism could be a plausible general theory of distributive justice that is also readily applicable to the problem of climate change.

References

Arneson, Richard. 2005. Distributive justice and basic capability equality: 'Good Enough' is not good enough. In *Capabilities equality: basic issues and problems*, ed. A. Kaufman, 17–43. London: Routledge.

Arrhenius, Gustaf, Jesper Ryberg, and Torbjörn Tännsjö (2022) "The Repugnant Conclusion", *The Stanford Encyclopedia of Philosophy* (Summer Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2022/entries/repugnant-conclusion/>.

Axelsen, D. V., & Nielsen, L. (2015). Sufficiency as freedom from duress. *Journal of Political Philosophy*, *23*(4), 406–426.

Benbaji, Y. (2005). The doctrine of sufficiency: A defence. *Utilitas*, *17*(3), 310–332.

[24] As noted above, some forms of prioritarianism allow aggregation in a way that will sometimes place the burdens on the worst off, if the morally weighted disutility of doing so outweighs the morally weighted disutility of placing it in the better off. Other versions (such as leximin) would give absolute priority to the worst off (among those above the threshold). Which version to choose, and whether to choose prioritarianism or egalitarianism are questions that fall beyond the scope of this paper.

Caney, S. (2005). Cosmopolitan justice, responsibility, and global climate change. *Leiden journal of international law*, *18*(4), 747–775.

Caney, S. (2010). Climate change and the duties of the advantaged. *Critical review of international social and political philosophy*, *13*(1), 203–228.

Caney, Simon (2014) Two Kinds of Climate Justice: Avoiding Harm and Sharing Burdens. *Journal of Political Philosophy* 22 (2): 125–149.

Casal, P. (2007). Why sufficiency is not enough. *Ethics*, *117*(2), 296–326.

Crisp, R. (2003). Equality, priority, and compassion. *Ethics*, *113*(4), 745–763.

Frankfurt, Harry (1987) Equality as a Moral Ideal. *Ethics* 98 (1): 21–43.

Gosseries, A. (2004). Historical emissions and free-riding. *Ethical perspectives*, *11*(1), 36–60.

Holtug, N. (2010). *Persons, interests, and justice*. Oxford University Press.

Haydar, B., & Øverland, G. (2014). The normative implications of benefiting from injustice. *Journal of Applied Philosophy*, *31*(4), 349–362.

Herlitz, A. (2018). The indispensability of sufficientarianism. *Critical Review of International Social and Political Philosophy*.

Huseby, R. (2010). Sufficiency: Restated and defended. *Journal of Political Philosophy*, *18*(2), 178–197.

Huseby, R. (2020). Sufficiency and the threshold question. *The Journal of Ethics*, *24*(2), 207–223.

Jagers, S. C., & Duus-Otterström, G. (2008). Dual climate change responsibility: on moral divergences between mitigation and adaptation. *Environmental Politics*, *17*(4), 576–591.

Kanschik, P. (2015). Why sufficientarianism is not indifferent to taxation. *Kriterion–Journal of Philosophy*, *29*(2), 81–102.

Knight, C. (2022). Enough is too much: the excessiveness objection to sufficientarianism. *Economics & Philosophy*, *38*(2), 275–299.

Nielsen, L. (2019). What is wrong with sufficiency?. *Res Publica*, *25*(1), 21–38.

Nielsen, Lasse. "Sufficiency grounded as sufficiently free: a reply to Shlomi Segall." *Journal of Applied Philosophy* 33, no. 2 (2016): 202–216.

Page, E. A. (2008). Distributing the burdens of climate change. *Environmental Politics*, *17*(4), 556–575.

Robeyns, Ingrid. 2017. Having too much. Pp. 1–44 in Jack Knight and Melissa Schwartzberg (eds), *NOMOS LVI: Wealth.* New York: New York University Press.

Segall, S. (2016). What is the Point of Sufficiency? *Journal of Applied Philosophy*, *33*(1), 36–52.

Shields, L. (2012). The prospects for sufficientarianism. *Utilitas*, *24*(1), 101–117.

Shields, L. (2016). *Just enough: Sufficiency as a demand of justice*. Edinburgh University Press.

Scanlon, T. (2018). *Why does inequality matter?*. Oxford University Press.

Timmer, D. (2022). Weighted sufficientarianisms: Carl Knight on the excessiveness objection. *Economics & Philosophy*, 1–13.

Widerquist, K. (2010). How the sufficiency minimum becomes a social maximum. *Utilitas*, *22*(4), 474–480.

Edward Page[1]

Benefiting at the Expense of Climate Change

> *'For this by nature is equitable, that no one be made richer through another's loss.'*[2]
> *'One cleans another's shoes; what can the other do but put them on?'*[3]

This paper discusses the problem of what to do, if anything, about the profits of activities that drive climate change. Should benefits created 'at the expense of' climate change be 'disgorged' to those who missed out and now face the adverse costs of the activities from which these benefits were created? The paper sets out to clarify the basis for disgorgement duties in private law and normative ethics and, in doing so, distinguishes between 'unjust enrichment' and 'wrongful enrichment.' It argues that the existence of the two tracks of unjustified enrichment is an established insight in the legal and ethical theory, but the significance of the distinction has yet to be fully explored in climate change justice. It is argued that neither approach generates a plausible case for legal recovery of unjust enrichments arising from climate change, but the wrongful enrichment track nonetheless serves as the basis of a powerful normative account of duties to disgorge profitable exploitations of the atmospheric commons.

[1] Edward Page, Department of Politics and International Studies, Warwick University, Coventry, UK. E.a.page@warwick.ac.uk

[2] Pomponius, cited in Gergen 2001: 1927.

[3] Baron Pollock, cited in Ripstein 2007: 1994n.

1. Introduction

Global climate change will impose significant costs on existing and future generations whatever measures are undertaken for its management.[4] Much of the normative literature on climate change has therefore addressed the important question of how these costs should be distributed amongst populations with contrasting needs, interests and responsibilities. What I want to focus on in this paper, however, is the 'benefiting side of the equation.'[5] What should be done, if anything, about the profits of activities that change the Earth's climate? According to one view, profits of injustice should be 'disgorged' to the victims of this injustice. [6] Could it be that those who profit from activities that drive climate change have a moral duty to disgorge this profit to the victims of unjustly imposed changes in climate? [7]

In this paper, I draw upon recent work in private law and normative ethics in order to explore disgorgement duties in the context of global climate change.[8] The paper seeks to make four main contributions. First, I clarify the structure and content of claims of unjustified enrichment in private law and to show that such claims can be developed in two senses (or 'tracks'): 'unjust enrichment' and 'wrongful enrichment.' The existence of these two tracks is an established insight in the literature on restitution in private law and normative ethics[9] but I hope to show that scholars have yet to explore its true significance for applied ethical debates such as global climate change. Second, I set out to show that neither track generates a plausible case in private law for recovery of unjust enrichments arising from climate change. Third, I argue that 'wrongful enrichment' may serve as the plausible starting point for a normative account of disgorgement duties in the context of climate change that is nonetheless independent of private law in justificatory terms. Fourth, I defend this account – which I call 'immoral enrichment' – from four objections.

---

[4] Greatly simplified, these costs are a combination of the cost of unprevented climate change combined with the cost of any measures undertaken to adapt to, or mitigate, climate changes that were regarded by policymakers as preventable.

[5] This phrase is taken from Wonnell 1996: 154.

[6] See Haydar and Øverland 2014; Goodin and Barry 2014; Butt 2009, 2014; Barry and Kirby 2017.

[7] Duties to disgorge benefits made at the expense of climate change have been proposed by Heyd 2017, Gosseries 2004, Page 2012, Lawford-Smith, 201, and Heyd 2017. Critics have responded equally forcefully, see Heyward 2014, Huseby 2015, 2018; Knight 2013; Lippert-Rasmussen 2017, 2022; Lindstad 2019.

[8] In what follows, I will refer to 'plaintiffs and defendants' (when discussing the law of unjustified enrichment) and 'victims, perpetrators and beneficiaries' (when discussing the normative ethics of unjustified enrichment).

[9] See Page 2012; Heyd 2017; Truccone-Borgogno 2023; Gilboa, Kaplan and Sarel 2024.

2. Unjustly benefiting from climate change

In private law, unjust enrichment involves 'an enrichment at the expense of another that has to be given up to that other for a reason, that reason being neither a contract nor a wrong.'[10] Assets are transferred from a plaintiff to a defendant in a way that renders the transaction defective and so reversible through a 'disgorgement' by the defendant directed to the plaintiff.[11] A paradigmatic example in many legal systems is a payment made in error, such as paying the same debt twice, but a wide range of other instances arise such as when emergency treatment is provided to someone who can only pay for this benefit at a later date or when a higher court reverses a lower court's enrichment from the funds of another. The core idea is that, where assets are defectively transferred from plaintiff to defendant it would be unjust for the defendant to resist disgorging the gain back to the plaintiff. Since the transfers of value in question do not involve a breach of a primary legal duty, the basis for disgorgement is merely 'the receiver was not entitled to it, nor intended to have it.'[12]

Unjust enrichment, as outlined above, can be analysed in terms of a five stage process: (1) a defendant is enriched (2) at the claimant's expense (3) in a manner that is unjust (4) thereby imposing a legal duty on the defendant to disgorge the benefit back to the plaintiff (5) subject to defences that limit their legal liability.[13] Below, I explain each step and explore how an unjust enrichment claim could conceivably be extended to climate change.

Enrichment. An enrichment in private law is usually understood as a 'restorable transfer of value'[14] that produces a 'favourable effect' on the recipient's interests.[15] The defendants in unjust enrichment are typically the initial, and direct, recipients of transfers of resources that the plaintiff believes should not have occurred or should not have occurred without a fair exchange of value. Usually, there is no question of a third party (or 'secondary defendant'[16]) acquiring a duty to disgorge but unjust enrichment law can be extended to indirect (or 'secondary') beneficiaries if their gains arose from integral features of the original unjustified transfer of value[17]

[10] Birks 2002: 497.

[11] In what follows, I will refer to 'plaintiffs and defendants' (when discussing the law of unjustified enrichment) and 'victims, perpetrators and beneficiaries' (when discussing the normative ethics of unjustified enrichment).

[12] Parke B, quoted in Birks 2005: 6. See also Birks (2001: 1789).

[13] See Birks 2001: 1791-93; Birks 2005:39-40; Barker 2008: 60–3; Burrows 2002: 15–51.

[14] Weinrib 2010: 655. See also Gergen (2001:1945n) and Birks (2005: 50-5).

[15] Gergen 2001: 1945–6.

[16] Häcker 2015: 50.

[17] Consider the case where D^1 (a bank) transfers a sum of money from P to D^2 by mistake.

or if the gains made by an immediate beneficiary were passed on intact to the current beneficiary after the fact.[18]

Turning to climate change, a climatic enrichment can be seen as a restorable transfer of value that owes its existence to climate change. Typically, the literature has focused on enrichments with origins in the acts and policies that drive changes in climate, but it is worth noting that a climatic enrichment could conceivably be created by a change in climate or from an act or policy that attempts to control climate change. What is required is to show that there has been a transfer of value between two parties mediated by climate change. Compared to simple cases of enrichments created by mistaken bank transfers, unpaid emergency services, or court mandated transfers of funds later ruled invalid, it is difficult to isolate the enrichments that different sorts of agents might not currently enjoy 'but for' climate change. But it is tolerably clear that huge benefits have arisen since the beginning of the industrial revolution that could not have arisen without climate change arising as a by-product. Although much of the total historical gain from climate changing activities may have been lost through consumption or waste – and much benefit may be so widely dispersed that a specific set of unjustly enriched defendants is difficult to identify – it is also tolerably clear that huge profits have been made by legal persons, such as large corporations, that have business models that rely on fossil fuels as their primary energy source. Individuals, households, and small and medium sized enterprises, by contrast, may be less suited as potential defendants since the benefits they derive will not stand out from other agents sufficiently for them to be named in court as an unjust beneficiary of climate change.[19]

At the expense of. In its most basic form, to benefit 'at the expense of another' involves one agent gaining 'from' another which usually involves a discernible transaction ('a nexus of exchange') through which wealth is transferred.[20] A transfer, in its most basic sense, is 'any action between persons.' The plaintiff must suffer some disruption (or 'normative loss'[21]) in their transaction with the defendant for a transaction to be said to be creating a benefit at the plaintiff's expense, but the plaintiff need not be made worse off materially through an unjustly enriching transaction. It is the unjust gain that is the focus of the plaintiff's demand for recovery and not any unjust losses they have experienced.[22]

[18] Consider the case where P drops his wallet by accident in the park. Later that morning D^1 picks it up and then the next day gives it later to his son, D^2, as a present.

[19] See Heyd 2017: 37; Weinbaum 2011: 450; Truccone-Borgogno 2023: 208; Gilboa, Kaplan, and Sarel 2024: 43.

[20] Birks 2005: 74-5; Smith, L. 2001: 2161.

[21] See Weinrib 1994: 283-4; Smith, L. 2001: 2141; Smith, S. 2001: 2188–90.

[22] Edelman and Bant 2016: 92; Smith, L. 2001: 2141.

Turning to climate change, if we can find a legal person that has profited distinctively and disproportionately from the activities that drive climate change, then it might be obvious that these profits were received 'as the expense of climate change' but, of course, since 'climate change' is not the plaintiff, the real question is at whose expense (if anyone's) have benefits from climate change been made? This is a far trickier question than I think has been recognized in the literature, but there seem to be three promising explanations. I do not think any of these is decisive but, together, they suggest that the critical question is whether these benefits are unjust and not whether they were made 'at the expense of' a potential plaintiff.

First, a victim of an impact of climate change residing in any nation or generation might link this experience to the profits of others by pointing out that those profits could not have been made without triggering the source of their misery.[23] If I face the negative consequences of your profitable activities then it does seem intuitive to say that 'I paid for your gain!' There is a doubt here that this causal ('but for') linkage of a benefit and a harm is strong enough to pass the 'at the expense of' test of private law but, in support of the idea, this test is not one of justice versus injustice but rather of establishing a basic sense of an enriching transfer obtaining between victim and beneficiary.

Second, we might think that these benefits were intercepted in a more direct sense from the assets of others. David Heyd, for example, argues that unjust enrichments have accrued as a result of the unequal use of the finite capacity of the atmosphere to absorb greenhouse gases. If everyone has an equal claim to the value of the greenhouse sink capacity of the atmosphere then those who profit from its unauthorised overuse can be seen as the beneficiaries of a mistaken transfer of value from those who have not benefited (or benefited less) from its exploitation.[24] It is worth noting that the accumulated gains of climate change are indeed distributed highly unevenly so it cannot be reasonably maintained that the plaintiffs and defendants in an unjust climatic enrichment analysis are one and the same. This means that it cannot reasonably be claimed that those benefits were not made at anyone's expense because all have equally benefited from climate change at each other's expense.

Third, and most radically, the defendants might be members of current generations who enjoy the valuable resource of climate stability that has been transferred to them from members of future generations. The idea is that large corporations and states operating in earlier generations are making windfall profits from activities that degrade climate stability and this amounts to a non-consensual transfer of

[23] Weinbaum 2011: 450.

[24] Heyd 2017: 38. See also Duus-Otterström 2014: 458; Page 2012: 315-6; Truccone-Borgogno (2022: 204–5).

value from the future to the present. Gilboa, Kaplan, and Sarel (2024: 42) model this account of 'at the expense of' on the valid enrichment of the defendant by a lower court that is later reversed by a higher court. This is an innovative solution but more intuitive is probably Heyd's interceptive enrichment idea which appeals to the same basic idea of the atmosphere being co-owned but lacks the somewhat strained appeal to court decisions that never occurred.[25]

The enrichment is unjust. It is vital to separate permissible from impermissible enrichments to avoid an implausible, and unworkable, account of the law of unjust enrichment. The paradigmatic example of a permissible benefit gained at the expense of another is a 'by benefit.' As Klimchuk (2007: 815) writes: 'If you live in the apartment above me, then, owing to the fact that heat rises, you will be enriched at my expense if I keep my apartment well heated through the winter. But you are not unjustly enriched.' Legal theorists are divided on how to separate unjust enrichments from 'by-benefits' but they are largely in agreement that there are two distinct, but generally converging, methods of doing so.[26] The first approach is to establish the presence of an 'unjust factor' that explains why an unjustly enriching transaction is defective in a way a by-benefiting transaction is not.[27] Such factors are 'all the possible matters between the plaintiff and defendant by which the plaintiff's intention to make a transfer is imperfect.'[28] In other words, the transaction between victim and beneficiary was legally impaired in a way that would justify its reversal that would not arise for a by-benefit.[29] The second approach is to determine whether or not the current holder can show why they have a legal basis to retain the enrichment and thereby resist the demand for restitution mounted by the plaintiff. Although the two approaches are based on contrasting rationales (the former starts on the basis that the enriching transfer was justified and attempts to impugn it, the latter starts on the basis that the enriching transfer is unjust until it can be justified), they are probably best seen as complements.[30]

---

[25] A separate problem with appealing to the legal rights of members of future generations to recover profits associated with past destabilisation of climate stability is the non-identity problem. Had these profitable activities not occurred then many, if not all, of the potential pool of defendants would likely never have existed. So it seems that arguing on behalf of future generations that current firms and states should not profit from exploiting existing climate stability at the expense of future individuals makes little sense. I believe that this problem can be solved if we assume that people coming into existence have a right not to be born in a state impaired by lack of access to a stable climate, but this is not a solution open to a private law approach given that the relevant rights do not yet exist in any legal sense that would make their future holders valid plaintiffs in a present-day court of law. See Caney 2006:474-6 and Page 2012: 319–20 for further discussion.

[26] See, for example, Birks 2005: 102-8; Klimchuk 2004: 1262–4.

[27] Smith 2001, L: 2163.

[28] Edelman and Bant 2006: 138 – original emphasis.

[29] See McBride and McGrath 1995: 36–7; Smith(S) 2001: 2122.

[30] See Klimchuk 2004: 1264; Edelman and Bant 2016: 130.

In the climate change context, there are at least three 'unjust factors' that have been explored in the literature.[31] I will argue that none of these factors corresponds to established bases of recovery of gains for the plaintiff[32] and all of them raise internal problems of coherence that question their use as legal grounds for recovery.

First, we might think that current agents enjoying climatic benefits should give up these benefits to compensate the victims who are harmed by the process(es) through which the benefits were created. All, or nearly all, people may end up being harmed by climate change but some benefit so much that their net losses are eliminated while the net losses of others are left in place. We may think it unfair to let the latter suffer while the former are allowed to profit. The problem with this line of reasoning is essentially the mirror image of the 'by benefits' problem noted above. Those gaining innocently from processes that cause harm may have gained 'at the expense of' others but they have not been unjustly enriched merely because their fortuitous gain is causally linked to the victims' impoverishment. If the defendants, meanwhile, can be held legally accountable for harming the plaintiffs, then this would transform the claim into one of compensation for wrongful harm and not restitution for unjust enrichment. Put differently, some may gain from processes that do not enrich (and may actively harm) others, but the former do not receive a transfer of value from the latter just because they gained more, or were harmed less, from a common activity.

Second, the injustice of retaining climatic benefits might be based on some populations involuntarily missing out on a fair share of benefits linked to the use of the capacity of the atmosphere to absorb greenhouse gas.[33] This is the corollary of Heyd's description of how agents benefit at the expense of others when they profit from their use of the atmospheric commons while bearing little or none of the associated costs imposed on others. Large corporations, for example, have boosted their profits by using fossil fuel energy sources at the cost of running up a debt of restitution to other atmospheric users whose own exploitation of the sink capacity of the atmosphere is now highly constrained. Of course, the exploitation of the resource cannot be 'given up' but the value created can be disgorged so that all benefit from the exploitation. Although I think this idea is broadly correct, the problem in this context is that it is only plausible as an account of wrongful, not unjust enrichment, since, if it does not appeal to wrongdoing in the production of the benefit, it is effectively an appeal to fairness or solidarity that does not supply a legal reason why this was a defective transfer of value at the time the transfer(s) occurred. The law of unjust enrichment, in this sense, is rigidly corrective in being

[31] See, for example, Truccone-Borgogno 2023: 205; Heyd 2017; Page 2012: 315–6.

[32] See Heyward (2014: 418n).

[33] See Heyd 2017: 38–9 and Page 2012: 315.

concerned with major disruptions in legal relationships caused by error and not about putting new arrangements in place. It may be the case that if there were a global legal rule in place that guaranteed all an equal share of the value of the absorptive capacity of the atmosphere then the profits of using more than your fair share should be redistributed. But this is not currently the case and the plausible claim that it is a mistake that this legal rule does not yet exist cannot deliver the conclusion that the imbalance of benefits derived from the resource arises from a mistaken transfer of value that legal authorities must reverse.

Third, the unjust factor may be that especially prosperous states and corporations operating earlier in history profit from activities that degrade climate stability at the expense of members of later generations who will inherit neither the climate stability of the present nor the benefit earlier gained from its degradation (Gilboa, Kaplan, and Sarel (2024: 42-3)). The idea is that the benefits gained by large corporations in the present amount to windfalls created by the mistaken decision of current legal authorities not to prohibit activities that threaten the birth-right of each generation to a stable climate. Much of the damage is already done but the element of this egregious intergenerational transfer of value concerning the profit of degrading climate stability can still be recovered by reversing the flow of profit that the lack of regulation of greenhouse gas emissions permitted. The problem with this ingenious account is that, even putting aside the problem of conceiving the unjust enrichment involved as if it were a valid but mistaken intergenerational transfer of value, the purported injustice done by profiteering corporations does not fit the logic of a claim of unjust enrichment. First, the plaintiffs do not yet exist and have no direct relationship with the defendants that could be the basis of a claim in private law. Second, since the identities of members of future generations are not fixed, but rather highly sensitive on actions and events leading up to their conception. It makes little sense to talk about a fiduciary making a claim on behalf of a plaintiff whose unjust windfall was a necessary condition of the defendant coming into existence. Third, the 'windfall' gained by the prospective defendants arises fundamentally because of the passage of time. Exploiting current climate stability for profit does not disrupt any current property rights that might be sensitive to changes in climate stability; and it is unclear what unjust factor would explain why the defendant ought not to retain the profit it makes from being lucky enough to operate in an earlier period in history where climate stability still obtains. Fourth, the claim seems not to be one of unjust enrichment at all. Instead, the idea seems to be that existing legal persons of a certain size are profiting disproportionately by wrongfully exploiting a resource in an analogous way to an agent benefiting from an unintended, yet highly profitable, trespass. The problem with this reasoning is not its internal coherence but rather, to borrow Birks (2002: 497) phrase, it invokes a

gain at the expense of a wrong rather than a gain from a defective transfer of value that it would be unjust not to reverse.

Remedy. Demonstrating that an enrichment arose in a transaction between plaintiff and defendant at the former's expense in an unjust manner does not specify a remedy for correcting this injustice. In the absence of an argument for a specific remedy, it could be that the unjust enrichment should be left where it lies. The remedy accepted by most legal theorists, and applied by case law in several countries, is that unjust enrichments should be restored ('given back') to the plaintiff. Although we cannot recreate the world as it would have been had the unjustly enriching transaction never happened, the plaintiff disgorging the gain to the victim makes it 'as if' the injustice between them had never arisen by making its effects the same as they would have been had the transaction been consensual. The defendant, who has done no wrong, is not liable for compensation for any loss endured by the plaintiff but they are liable to return what was the plaintiff's back to them. This is a two-step process: first, they disgorge and, second, that which is disgorged is directed to the plaintiff.[34] The exact nature of the disgorgement may vary with the account given of enrichment, at the expense of, and injustice in that case, but a common assumption is that the defendant must disgorge up to the point where they no longer benefit from the plaintiff and this means that they may be compensated for any costs they incurred in receiving, holding or improving the benefit.[35] What unjust enrichment does not do is scale the action required of defendants to any direct costs, or harms, associated with the activities that generated the unjust benefit and so responding to the harms imposed by climate change lies outside the claim. This reflects the structure and justification of unjust enrichment as one of corrective, and not distributive, justice. Correcting unjust climate enrichment will involve the defendants giving up to plaintiffs the total value of the former's enrichment at the latter's expense minus costs incurred and not the total value required to compensate the latter for loss and damage caused by climate change.

Defences. The above steps are integral parts of the cause of action in unjust enrichment. If navigated successfully, a prima facie liability arises on the part of a defendant to return the value of an unjust enrichment to the plaintiff. The final step concerns the assessment of any defences that reduce the liability of defendants to provide restitution. This strongly differentiates the last stage of unjust enrichment from its predecessors since showing that the defendant was in fact not enriched – or it was not at the expense of (or unjust to) the plaintiff – dispenses altogether with the need for the beneficiary to offer a defence against disgorgement. There are two

[34] Birks 2005: 17–18.

[35] Ripstein 2007: 19943–4.

such defences available to a defendant seeking to resist a restitution of unjust enrichment in the climate context. First, 'change of position' is a defence where the defendant argues that, through no fault of their own, they are no longer in possession of the enrichment made at the expense of the plaintiff and hence any disgorgement they might otherwise have owed is nullified. The manner of this disenrichment must have been a lawful activity (whether a holiday, a gift, or some other service) that would not have occurred 'but for' the defendant being enriched by the plaintiff.[36] The logic of the change of position defence is that reversing their enrichment to the benefit of the plaintiff cannot make the defendant worse off than they would have been had they not been unjustly enriched. Given that many unjust climatic enrichments will have accrued to agents who relied upon these enrichments in good faith to undertake lawful activities that they would not otherwise have undertaken, this defence seems very wide-ranging in the climate case. Second, 'bone fide exchange of value' is a defence where the defendant argues that, since they paid fair value for the good alleged to have been unjustifiably transferred to them from the plaintiff, they do not exist in a state of unjust enrichment that needs to be reversed through a disgorgement to the plaintiff. This defence is important since, while the original beneficiary of a defective transfer might not have access to this defence since they indeed received something for nothing, any subsequent recipient of a climatic enrichment may have exchanged fair value for it and so will not be duty bound to disgorge the enrichment. This suggests that making a case for disgorgement of profits accumulated in industries with many steps will be very difficult indeed since each subsequent beneficiary will have greater access to the 'exchange of value' defence.[37]

3. Wrongful enrichment

As we have seen, the 'unjust' track of unjustified enrichment can conceivably be applied to climate change but it is weakened considerably by problems of identifying the unjust factor in the creation of climate change benefits and the strength of defences that defendants would likely be able to mount against disgorgement. This gives us reason to consider the alternative, 'wrong', sense of unjustified enrichment. In private law, wrongful enrichment (or 'enrichment by wrong'[38]) focuses on the recovery of gains arising from profitable breaches of duty.[39] Unlike unjust enrichment, it is the benefit creating features of the wrong, and not the unjustified transfer

[36] See Birks 2001: 1786-87; Birks 2005: 207-64; Edelman and Bant 2016: 332–48.

[37] See Birks 2002: 525.

[38] Klimchuk 2004: 1261. See also Birks 2001: 1782.

[39] See Birks 2001: 1783-86; Burrows 2002: 455–62; Klimchuk 2004: 1259–61; Wonnell 1996: 160–1;

of assets from plaintiff to defendant, that is at the heart of the plaintiff's cause of action.[40] A paradigmatic example of wrongful enrichment is when a defendant uses without authorisation the plaintiff's property or person for profit making purposes that do not harm the plaintiff materially. In these cases, the wrongful gains of the defendant are not generated directly by a defective transfer of assets between the two parties, so there has been no obvious gain made 'at the victim's expense.' Instead, the gain has been made at the expense of a breach of duty owed to the victim.

A claim of wrongful enrichment is, in the above respects, more conceptually economical, but also more demanding, than unjust enrichment. It is more economical because identifying a profitable wrong replaces the more complex process of identifying an enrichment and then explaining how this was made at the plaintiff's expense such that it would be unjust not to reverse. All that is required in wrongful enrichment is to demonstrate that a gain is a product of a wrong committed by the defendant against the plaintiff and to combine this fact with a core normative principle of corrective justice that wrongs should be defeated as a second best for them not being violated in the first place.[41] It is more demanding in two ways. First, a breach of legal duty must be identified, and this excludes all forms of unjustified enrichment that do not involve wrongdoing. Second, the wrongly enriched, on most accounts, have more extensive duties of disgorgement than the unjustly enriched since they are not reversing a faulty transfer of value but rather giving up the value of all of their profit despite this never being held by the plaintiff.[42]

The stages of the wrongful enrichment claim can thus be summarized as follows: (1) there is an enrichment on the part of the defendant (2) arising from a breach of duty owed to the plaintiff (3) that creates a liability on the part of the defendant to surrender the enrichment in favour of the plaintiff (4) subject to defences that mitigate this liability.

Enrichment. An enriching wrong must involve a material benefit arising for a defendant that can, in principle, be taken from them for the purpose of defeating the wrongdoing through which it was created. The benefit must be directly (that is, immediately) connected to the wrong in the sense that it was an intended – or, if not intended, then a constitutive element – of the breach of duty committed against the plaintiff.[43] Wrongful enrichments may involve money or other assets that are taken, exploited, or sold without the permission of the owner. They could also involve profits gained from exploiting or selling assets where a legal duty owed to the

McInnes 2015: 250–2.

40 Birks 2005: 74; Burrows 2002: 455; Virgo 2006: 425–8.

41 See Gergen 2001: 1931; Smith 2001: 2116.

42 See Gergen 2001: 1933–38.

[43] Virgo 2006: 448–9.

plaintiff restricts the exploitation of these assets. Wrongful enrichments generally only arise for the immediate beneficiaries of wrongdoing and functionally become unjust enrichments when they are passed on to subsequent beneficiaries.

In terms of climate change, wrongful and unjust enrichment raise similar issues of application with the common point being that much of the wealth created by successive generations of atmospheric users would not have arisen 'but for' the activities that drive climate change. The difference in the analysis is whether it must be shown that these enrichments are generated from impaired transfers and exploitations of assets owned by the plaintiff (unjust enrichment) or from wrongs committed against the plaintiff (wrongful enrichment). An enrichment could arise from a direct profiting from wrongdoing (as when a firm makes a financial gain from violating its legal duty to mitigate or adapt to climate change) or from an indirect profiting from wrongdoing (as when a firm makes a financial gain from a transaction made possible by another firm profiting from violating its legal duty to mitigate or adapt to climate change). As with indirect unjust enrichment, the move from a direct to indirect wrongful enrichment may or may not be a more promising fit for the climate problem, all things considered, since the enrichments in question may or may not be best conceived as the accumulated gains of wrongdoers or those that do profitable business with wrongdoers.

Wrongdoing. Showing a breach of a legal or moral duty has occurred effectively replaces two stages in the unjust enrichment framework because it eliminates the need to show an enrichment arose (1) at the expense of the plaintiff and (2) an unjust factor (such as impaired consent) was present in the creation of the enrichment. In the wrongful enrichment framework, it is the connection between wrongdoing and a gain from this wrongdoing which triggers a claim for restitution. Typical examples of enriching wrongs are when someone deceives another into transferring money to them, or when someone is paid to assault someone, or when someone sets out to profit from another person's image or property without consent. All these cases clearly involve an enrichment (money) gained at the expense of an intentional wrong committed against another agent. But they are not best understood as defective transfers of value from plaintiff to defendant since the benefits concerned were never in the possession of the plaintiff.

In the climate change context, wrongful enrichment only covers gains from climate changing activities that were in breach of a legal duty and this seems to reduce the scope of the account considerably since it is obvious that the profits of most activities that drive climate change did not originate in a breach of any specific legal duty.[44] Some response to this concern arises from the consideration that the

[44] Duus-Otterström 2014: 458.

breaches of legal duty are not limited to intentional breaches of environmental regulations but may also be unintentional breaches of customary law (such as profitable trespasses). So the wrongful enrichment track may still have considerable scope if it can be shown that profits have been made in breach of such norms. Gilboa, Kaplan and Sarel (2024: 39-40) usefully list three avenues of wrongful climatic enrichment that could meet the breach-of-legal-duty test: profitable violations of an 'explicit environmental regulation' (e.g. gains made from deliberately exceeding a legal emissions limit), acting in a 'grossly unreasonable or negligent' manner (e.g. gains made from carelessly emitting more greenhouse gas than was needed for the activity concerned), and profiting from corrupt behaviour (e.g. hiding environmental impacts through falsifying environmental performance). However, despite the existence of duties matching these three duty types in many states – and the evolving UNFCCC legal architecture where developed states have agreed to act 'as if' they were bound by legal mitigation duties – the scope of legal duties that would ground claims of wrongful enrichment is very limited indeed. Nevertheless, the prospect of more explicit legal norms against climatic enrichment in the future means that the breach-of-legal-duty test may become more easily met in the future even though plaintiffs may struggle to identify specific enrichments that would not exist 'but for' the breach of legal duty of the defendant.

Remedy. The remedy specified by wrongful enrichment is subtly different to that of unjust enrichment and may result in quite different remedial demands being made on beneficiaries. This is due to the presence of the tortious, or wrongful, element in the transaction that creates the defendant's enrichment. We can say that the approaches diverge in two main respects in terms of the remedies recommended. First, since there are many types of legal wrongs, the question must be asked what remedy is required to correct the injustice between the parties. Some profitable wrongs (such as profitable but unintentional trespasses) may require less extensive restitution than other profitable wrongs (such as intentional misuse of intellectual property) since making it as if the wrongful enrichment never happened (the primary purpose of recovery of wrongful enrichments) may require less extensive restitution in the former case because the breach of duty to the plaintiff is less serious.[45] Second, and more importantly, it is natural to conceive the correction of the forms of unjustified enrichment as diverging in their general approach to how defendants can make their unjustified enrichments right again. Put simply, the wrongfully enriched should give up all of the profits of their wrongdoing whereas the unjustly enriched should give up only as much as would simulate the alternate world where the unjust transfer had never happened and this usually means deducting

[45] See Birks 2001: 1792–3; 1961–65; Rotherham 2007: 190–3; Burrows 2002: 461–2.

from the disgorgement a reasonable approximation of the defendant's costs in receiving the enrichment.[46] The idea is that, in restitution for wrongful enrichment, we are not merely reversing a transfer of value that should not have happened but also eliminating any trace of the defendant profiting from wrongdoing committed against the plaintiff.[47] Where this is relevant for climate change justice is that recovery of wrongful enrichments from breaches of legal duty would seem to reach beyond the initial value of the enrichment a state or firm gained from activities emitting greenhouse gas into the atmosphere to encompass all profits later derived from these activities.

Defences. In general, defences to liability in wrongful enrichment are far less extensive than those of unjust enrichment. The 'change of position' defence is not available to the most important category of wrongfully enriched (law breakers) since they cannot argue that they relied in good faith upon a wrongful enrichment to undertake an activity or discharge a debt; and it is only open in a limited way to the intended beneficiaries of wrongdoers who are in a similar position to those who unknowingly receive stolen goods.[48] Subsequent beneficiaries (those that benefit innocently from later transactions with the wrongdoers) will have access to additional defences. These secondary beneficiaries, who enjoy a considerable proportion of the profit of activities that drive climate change, will be able to appeal to a 'change of position' defence (an example might be shareholders of oil companies who relied upon dividends to pay debts) or a 'fair value' defence (an example might be contractors of oil companies that charged a market rate for corporate services). The wide scope for defences of this sort on behalf of secondary beneficiaries of climate change appears to restrict the usefulness of the private law of wrongful enrichment considerably in the climate context even if it could be shown that those who exploit the finitude of the atmospheric sink for profit violate a legal duty in so doing.

4. Immorally benefiting from climate change

To sum up the paper so far, both unjust enrichment and wrongful enrichment are promising approaches to the 'benefit side of the question' but the legal wrongdoing required is not generally present to ground claims of wrongful climatic enrichment and, whatever is unjust about profiting from activities that drive climate change, this is not readily explained in terms of a mistaken payment or court ordered

[46] Gilboa, Kaplan, and Sarel 2024: 34.

[47] For more on this, see Ripstein 2007.

[48] See Birks 2001: 1787. For a detailed discussion of defences to disgorgement of unjust and wrongful enrichments, see Edelman and Bant 2016: 363-403.

transfer subsequently reversed. This does not necessarily mean we need to give up on unjustified enrichment as being a part of climate change justice since this doctrine can be developed as a normative doctrine that lacks the restrictions, and is justified independently, of private law. According to the proposed normative account, which I call 'immoral climatic enrichment', moral agents acquire disgorgement duties when they benefit from wrongdoing in the way the absorptive capacity of the atmosphere is exploited. The wrongdoing involved is essentially that large corporations and states profit disproportionately from unauthorised exploitations of the earth's capacity to absorb greenhouse gas without fully compensating those who have been excluded from these profits. This account is broadly compatible with the accounts of profitable exploitation of the atmospheric commons proposed by Page (2012), Heyd (2017) and Gilboa et al (2024) but is based on correcting violations of a moral duty not to profit from unilateral exploitations of resources that others rely upon to pursue their ends rather than the reversal of mistaken transfers of value in absence of wrongdoing.

The four stages of immoral enrichment

Benefit. The currency of immoral enrichment is similar to that of its legal corollaries – a restorable transfer of value equivalent to a cost saved or a debt paid. The benefit, in more concrete terms, is the financial value of any exploitation of the absorptive capacity of the atmosphere that was not shared fairly with other atmospheric users. Agents may immorally benefit from climate change in three ways. First, they may intentionally, or unintentionally, benefit disproportionately from their own climate changing behaviour ('direct enrichment'). Second, they may be the beneficiaries, intentionally or unintentionally, of the climate changing behaviour of another agent ('constitutive enrichment'). Third, they may benefit from an initial enrichment, in either of these two ways, after it is passed to them in a subsequent transaction ('sequential enrichment'). When so they benefit, subject to certain defences, the enriched should give up profits in favour of those excluded from these profits.

Wrongdoing. The benefits described above were made at the expense of a wrong committed against other agents who were excluded from sharing in the profitable exploitations of the absorptive capacity of the atmosphere that were not compatible with long-term climate stability. What made it wrong was the agents involved, best conceived as large states and corporations, continued to profit from excess greenhouse gas emissions even though they were aware of the science of climate change and endorsed the normative goal of limiting climate change. At some point in the recent past, it ceased to be reasonable to treat these agents as behaving non-wrongfully in respect of activities that made profits by degrading a valuable resource that

they ought not to treat as their own. Essentially, these agents exploit location at a fortuitously earlier moment of time where profits may still be legally internalized, and costs externalized, through use of the atmospheric sink due to lack of regulation that is widely known to be necessary to protect the climate system. This explains the lack of a legal wrong in unjust climatic enrichment since the excess profit taking involved is enabled by the lack of legal regulation and not in violation of the regulations that do exist. Nevertheless, there is still an injustice present in such behaviour that we can correct by defeating the wrongdoing associated with exploiting the lack of legal regulation currently protecting use of the atmospheric commons for profit. Put slightly differently, valuable means have been intercepted from others who have been excluded from the profits made from their means. Whilst the degradation of the atmospheric sink might have been unintended, the creation of profit from the activities that degraded the atmosphere was intentional. In this sense, the wrongdoing involved is analogous to profiting from selling a dwindling supply of drinking water from a lake located on public property that one is aware is relied upon by others who had equal access rights but is not regulated in any other way.

Remedy. The remedy for immoral enrichment is the corrective one of making it as if the injustice between the parties to the enrichment had never happened which can also be seen as the state of affairs where wrongdoing is no longer present in that transaction.[49] This is achieved by simulating the alternate reality where all parties had benefited as they should have from profitable use of the means to which they had equal claims. It is, in another description, to transform the immorally enriched into an agent of all through disgorgement so that the immoral enrichment never happened in the sense that no one is now wrongfully enriched. This is achieved by disgorging the profit to all moral agents that can claim to have an equal right to use and profit from the atmospheric sink, which would more than likely include the unjustly enriched as well as those excluded from any profit. It is probably most useful to imagine this correction being carried out by an international restitution scheme funded by a windfall tax on large corporations. This would dispense with the costs and difficulties of court-based recovery and the billions of potential plaintiffs and defendants who might be involved. The justification for the fund would be that it turns the situation into one where nobody does wrong in respect of profiting from climate change if the profits are disgorged to an international fund tasked with redistributing these profits as if they had been created non-wrongfully.

Defences. Much of the benefit created by activities that cause climate change has been consumed by the original beneficiaries. Other benefits were indirectly

[49] As Ripstein puts it, '[i]f I use what is yours, without your consent, I wrong you. The problem is coming up with the way in which that wrong can be righted, and the only way it can be righted is turning it into a situation in which nobody does wrong after all' (2007: 1994n).

received by agents who have transformed, or added to, these benefits in good faith. The immoral enrichment account, despite not being grounded in private law, inherits the idea that benefits like the above need not, or cannot, be enriched. Despite moral norms around the permissible exploitation of the atmospheric sink becoming clearer in recent decades, some agents may maintain they were excusably ignorant of the duty not to profit from degrading the atmospheric sink so profits accumulated in the past may lie beyond the account. Other beneficiaries of the depletion of the capacity of the atmosphere to absorb greenhouse gas may have relied upon their profits in good faith to pay off debts or discharge other legal duties that would not have arisen but for the activities from which they were enriched. It is worth noting, however, that the remaining profit will be substantial given the upper baseline for recovery is all future profit from activities that deplete the atmospheric sink until the climate system is returned to a safe equilibrium. So the account seems to have considerable scope despite the defences of 'change of position' or 'fair exchange of value.' Even if the reduction in liability is total, there is still the potential to require those profiting from climate change to undertake non-material actions of restitution familiar to accounts of transitional justice such as apologies or participation in truth and reconciliation processes.

Four objections

I have space, here, to consider four objections to the immoral enrichment account which are finely balanced between objections to the way the account justifies disgorgement duties in any context and objections to the way the account has been applied to climate change.

First, one might not contest the validity of the immoral enrichment account, especially in small scale cases, but instead contest its value as an approach to the problem of climate change justice. The account, as with its legal analogues, rests on a sharp separation between enrichment ethics and impoverishment ethics that may seem unsettling if the hope was for the account to take a leading role in the task of justly allocating the costs of responding to climate change. In response, it is worth noting the immoral enrichment account is compatible with more distributive accounts (concerned with allocating costs of responding to climate change) as well as torts (concerned with compensating for wrongful losses). Moreover, those who secure restitution for immoral climatic enrichment – perhaps in the form of payouts from a global restitution fund – would have a reasonable complaint that it would be unjust if these restitutions were made on condition that they replace sources of finance for tackling climate change.

Second, the immoral enrichment account requires wrongdoing on the part of the

agent that initiates an enrichment through this wrongdoing. However, the account also includes within the scope of disgorgement benefits held by agents who have benefited indirectly from wrongdoing. In the climate change context, this means the proposed disgorgement can reach benefits enjoyed by agents who have not changed the climate in any meaningful way or directly violated any primary moral duty not to profit from climate change. This seems to treat the innocent beneficiaries of wrongdoing as if they were the perpetrators of wrongdoing and this seems unfair. In response, the idea is that remote beneficiaries, although they have committed no breach of moral duty in the production of a wrongful benefit themselves, would be breaching a duty of corrective justice if they fail to play their special role in reversing the transaction that left them in possession of this benefit. As a non-wrongdoer, however, they have no duty to step into the shoes of the perpetrator; and not being party to the primary transaction between perpetrator and victim, they cannot reverse that transaction. They can, however, reverse one of that transaction's key unjust effects, namely, the creation of a benefit for them that is tainted through wrong done to the victim. They do this by transferring the enrichment to the victim.

Third, the inclusion of sequential immoral enrichments in the account might seem problematic for a further, conceptual, reason since it is a revision to standard accounts of corrective justice that typically assume there is only one correctible transaction at the heart of a corrective injustice: the transaction between victim (plaintiff) and perpetrator (defendant). In immoral enrichment, this is the transaction between someone who benefits from their own wrongdoing and the person who demands disgorgement of benefits made from the wrong done to them.[50] The objection, here, is essentially that, whereas it may make sense in the abstract to talk of a secondary transaction where the beneficiary is not the initial wrongdoer, this insight cannot be accommodated within the correlative structure of corrective justice and the mere possibility we can conceive of things this way does not generate any duty independent of corrective justice.[51] In response, corrective justice concerns the righting of transactions that it would be wrong to leave intact and a transaction is, in essence, 'any action between persons.'[52] The primary transaction in immoral enrichment (the wrong done by perpetrator to victim) clearly fits this understanding and it is a vital ingredient to a corrective analysis of what has gone wrong in immoral enrichment. Nevertheless, the secondary transaction that arises in cases of constitutive enrichment (the enriching of a beneficiary as a direct product of the wrong done to the victim) and sequential enrichment (the enriching of beneficiary by perpetrator after the wrongdoing) is no less of a transaction merely

[50] See Barker 1995: 469; Weinrib 2010.

[51] McBride 2015: 260-2; Smith, L. 2001.

[52] Edelman and Bant 2016: 92 – original emphasis.

because it is secondary. There is 'action' between the victims and beneficiaries in these latter cases that can be corrected. First, a benefit has been transferred between these parties in the sense that one agent is now in possession of value created by a wrong done to another. Mere possession of the wrongful benefit puts the beneficiary in a position where they can either assist in defeating the wrongdoing or contribute to its continuing existence. Second, if a sequential or constitutive beneficiary refuses to transfer the benefit to the victim, they in effect create a new wrongful transaction between themself and the victim since any further profit they make will be knowingly made from a wrong that was never corrected.

Fourth, in response to a common objection that corrective justice crowds out distributive justice, it is frequently responded that, in only seeking to reserve space for the reversal of wrongful gains and losses, corrective justice leaves virtually unlimited space for distributive interventions motivated by egalitarian and other ideals.[53] The problem arises that, if the corrective account is indeed limited to the reversal of very specific transactional failures, then the immoral enrichment account would be redundant. The objection may be put like this: since every restitution of immoral enrichment will be subsequently checked against the preferred pattern of distribution, it is surely this pattern of distribution and not the operation of corrective justice that will ever determine how people finally fare.[54] In the climate context, the idea is that any conceivable distributive solution will sweep away any limited redistributions of immoral gains between specific atmospheric users thus raising the question of what the point might be of identifying and enforcing disgorgement duties? In response, the application of many distributive ideals will leave the application of corrective justice intact if the realization of the preferred distributive pattern is not threatened by the correction in question. Much, of course, will depend on the ideal of distribution but those with close ties to human rights, basic needs, and maintaining decent lives, will all leave significant room for corrections of wrongdoings such as immoral enrichments. Suppose, however, that the distributive ideal endorsed did, in every case of immoral enrichment, leave the final distribution of entitlements with no trace of prior operations of corrective justice. It would be known that these corrections would have been made had they not been overruled by our distributive ideals and the parties to the flawed transaction would be aware the injustice to which they were linked was taken seriously as a corrective injustice. In such circumstances, it is not as if the correction had never happened but rather the correction reversed an injustice between parties that was subject to a further intervention on distributive grounds.[55]

[53] See Page and Duus-Otterström 2023: 20–1.

[54] See Lippert-Rasmussen 2017: 80–1; Knight 2013: 585–6; Parr 2016: 992.

[55] See Klimchuk 2003: 63–4.

5. Conclusion

In this paper, I drew upon recent work in ethics and private law to clarify the doctrine of unjustified enrichment, and explore how this doctrine might be applied to global climate change. I argued that neither of the two main tracks of unjustified enrichment law is promising as a justification for disgorgement duties in the climate context. I then argued that the idea of 'wrongful enrichment' could be adapted as an independent normative account in a way that addresses an important and neglected dimension of the injustice posed by climate change. This is that moral agents that continue to profit from climate changing activities with wrongful origins should, subject to defences, disgorge these profits to moral agents excluded from these benefits. I developed this argument more fully through an analysis of complex transactions of immoral climate enrichment and the four stages of an immoral climate enrichment claim. I then defended this account from four objections. The argument defended suggests that the near exclusive focus of international organisations on avoiding (or, if unavoidable, compensating for) losses and damages arising from anthropogenic climate change may have blinded us to an independent and no less egregious injustice: profiting from anthropogenic climate change in a way that wrongly excludes others.

References

Barker, K. (1995) 'Unjust Enrichment: Containing the Beast', Oxford Journal of Legal Studies 15: 457–75,

Barker, K. (2008) 'Responsibility For Gain: Unjust Factors or Absence of Legal Ground? Starting Points in Unjust Enrichment Law', in R. Grantham and C. Rickett, Structure and Justification in Private Law. Oxford: Hart, pp. 44–74.

Barry, C, & Kirby, R. (2017). Skepticism about Beneficiary Pays: A Critique. Journal of Applied Philosophy, 34(3), 285–300.

Birks, P. (2001) 'Unjust Enrichment and Wrongful Enrichment', Texas Law Review 79: 1767–94.

Birks, P. (2002) 'At the Expense of the Claimant: Direct and Indirect Enrichment in English Law', in D. Johnston and R. Zimmermann (eds) Unjustified Enrichment: Key Issues in Comparative Perspective. Oxford: Oxford University Press, pp. 493–25.

Birks, P. (2005) Unjust Enrichment. Second Edition. Oxford: Oxford University Press.

Burrows, A. (2002) The Law of Restitution. Second Edition. London: Butterworths.

Butt, D. (2009) Rectifying International Injustice. Oxford: Oxford University Press.

Butt, D. (2014) "A Doctrine Quite New and Altogether Untenable': Defending the Beneficiary Pays Principle', Journal of Applied Philosophy 31(4): 336–48.

Caney, S. (2006). Environmental Degradation, Reparations, and the Moral Significance of History. Journal of Social Philosophy, 37(3), 464–482.

Duus-Otterström, G. (2014). The Problem of Past Emissions and Intergenerational Debts. Critical Review of International Social and Political Philosophy, 17(4), 448–469.

Duus-Otterström, G. (2017) 'Benefiting from Injustice and the Common-Source Problem', Ethical Theory and Moral Practice 20 (5): 1067–81.

Duus-Otterström, G. and Page, E.A. (2023) 'Why Victims of Unjust Harm Should Take Priority over Victims of Bad Luck', Social Theory and Practice 49(1): 57–77.

Edelman, J. and E. Bant (2016) Unjust Enrichment. Second Edition. Oxford: Hart Publishing.

Gardner, J. (2011) 'Torts and Other Wrongs', Florida State University Law Review 39(43): 43–64.

Gergen, M. P. (2001) 'What Renders Enrichment Unjust', Texas Law Review 79: 1927–80.

Gilboa, M., Kaplan, Y and Sarel., R. (2024) 'Climate Change as Unjust Enrichment', July 6, Georgetown Law Journal, forthcoming.

Goodin, R. (2013) 'Disgorging the Fruits of Wrongdoing', American Political Science Review 107(3): 478–91.

Goodin, R.E. and Barry, C. (2014) 'Benefiting from the Wrongdoing of Others', Journal of Applied Philosophy 31(4): 363–75.

Gosseries, A. (2004) 'Historical Emissions and Free-Riding', Ethical Perspectives 11(1): 36–60.

Grantham, R.B. (2003) 'Disgorgement for Unjust Enrichment', Cambridge Law Journal 62(1): 159–80.

Häcker, B. (2015) 'Enrichment in Three-Party Cases', in C. Mitchell, & W. Swadling (Eds.), The Restatement Third: Restitution and Unjust Enrichment: Critical and Comparative Essays (pp. 31–57). Oxford: Hart.

Haydar, B. and Øverland, G. (2014) 'The Normative Implications of Benefiting From Injustice', Journal of Applied Philosophy 31(4): 349–62.

Heyd, D. (2017) 'Climate Ethics. Affirmative Action, and Unjust Enrichment', in L.H. Meyer and P. Sanklecha (eds) Climate Change and Historical Emissions. Cambridge: Cambridge University Press, pp. 22–45.

Heyward, J.C. (2014) Benefiting from Climate Geoengineering and Corresponding Remedial Duties: The Case of Unforeseeable Harms. Journal of Applied Philosophy, 31(4), 405–19.

Huseby, R. (2015) 'Should the beneficiaries pay?', Politics, Philosophy and Economics 14(2): 209–25.

Huseby, R. (2018) 'Should We Relinquish or Distribute the Benefits of Injustice', Journal of Value Inquiry 52: 213–225.

Klimchuk, D. (2003) 'On the Autonomy of Corrective Justice', Oxford Journal of Legal Studies 23(1): 49–64.

Klimchuk, D. (2004). Unjust Enrichment and Reparations for Slavery. Boston University Law Review, 84: 1257–1275.

Klimchuk, D. (2007) 'The Scope and Structure of Unjust Enrichment', University of Toronto Law Journal 57(4): 795–817.

Knight, C. (2013) 'Benefiting from Injustice and Brute Luck', Social Theory and Practice. 39(4): 581–98.

Lawford-Smith, H. (2014) 'Benefiting from Failures to Address Climate Change', Journal of Applied Philosophy 31(4): 392–404.

Lindstad, S. (2020). What Is the Wrong in Retaining Benefits from Wrongdoing? How Recent Attempts to Formulate a Plausible Rationale for the 'Beneficiary Pays Principle' Have Failed. Res Publica, 26(1), 25–43.

Lippert-Rasmussen, K. (2017) 'Affirmative Action, Historical Injustice, and the Concept of Beneficiaries, The Journal of Political Philosophy 25(1): 72–90.

McBride, N.J. (2015) 'Restitution for Wrongs', in C. Mitchell, & W. Swadling (Eds.), The Restatement Third: Restitution and Unjust Enrichment: Critical and Comparative Essays. Oxford: Hart, pp. 251–84.

McBride, N.J. and P. McGrath (1995) 'The Nature of Restitution', Oxford Journal of Legal Studies 33: 33–49.

McInnes, M. (2015) 'The Restatement Third: Restitution and Unjust Enrichment', Canadian Business Law Journal 57: 235–52.

Page, E. (2012) 'Give it up for climate change: a defence of the beneficiary pays principle', International Theory 4(2): 300–30.

Page, E.A. and Duus-Otterström, G. (2023) 'Correcting Unjust Enrichment: Explaining and Defending the Duty to Disgorge the Benefits of Wrongdoing', Critical Review of International Social and Political Philosophy. (in press).

Parr, T. (2016) 'The Moral Taintedness of Benefiting from Injustice', Ethical Theory and Moral Practice 19: 985–97.

Pogge, T. (2008). World Poverty and Human Rights. Second Edition. Cambridge: Polity.

Ripstein, A. (2007) 'As If It Never Happened', William and Mary Law Review 48: 1957–97.

Rotherham, C. (2007) 'The Conceptual Structure of Restitution for Wrongs', Cambridge Law Journal 66: 172–99.

Smith, L. (2001) 'Restitution: The Heart of Corrective Justice', Texas Law Review 79: 2115–2175.

Smith, S. (2001) 'Justifying the Law of Unjust Enrichment', Texas Law Review 79: 2177–97.

Truccone-Borgogno, S. (2023). Climate Justice and the Duty of Restitution. Moral Philosophy and Politics, 10(1), 203–224.

Virgo, G. (2006) The Principles of the Law of Restitution. Second Edition. Oxford: Oxford University Press.

Weinbaum, A. (2011) 'Unjust Enrichment: An Alternative to Tort Law and Human Rights in the Climate Change Context', Pacific Rim Law and Policy Journal 20: 429–54.

Weinrib, E.B. (1994) 'The Gains and Losses of Corrective Justice', Duke Law Journal 44: 277–97.

Weinrib, E.J. (2010) 'Unjust Enrichment', in D. Patterson (ed) A Companion to Philosophy of Law and Legal Theory. Second Edition. Oxford: Blackwell, pp. 654–65.

Wonnell, C.T. (1996) 'Replacing the Unitary Principle of Unjust Enrichment', Emory Law Journal 45: 153–220.

Studies on climate ethics and future generations, vol. 1 Working paper series 2019:1–11. Eds. Paul Bowman & Katharina Berndt Rasmussen

Tim Campbell: *The Bullet-Biting Response to the Non-Identity Problem*

Melinda A. Roberts: *Does the Additional Worth-Having Existence Make Things Better?*

Anders Herlitz: *Nondeterminacy and Population Ethics*

Wlodek Rabinowicz: *Can Parfit's Appeal to Incommensurabilities Block the Continuum Argument for the Repugnant Conclusion?*

Gustaf Arrhenius & Julia Mosquera: *Positive Egalitarianism*

Marc Fleurbaey & Stéphane Zuber*: Discounting and Intergenerational Ethics*

Stéphane Zuber: *Population-Adjusted Egalitarianism*

Katie Steele: *'International Paretianism' and the Question of 'Feasible' Climate Solutions*

Göran Duus-Otterström: *Sovereign States in the Greenhouse: Does Jurisdiction Speak against Consumption-Based Emissions Accounting?*

Paul Bowman: *On the Alleged Insufficiency of the Polluter Pays Principle*

Martin Kolk: *Demographic Theory and Population Ethics – Relationships between Population Size and Population Growth*

Studies on climate ethics and future generations, vol. 2 Working paper series 2020:1–11. Eds. Paul Bowman & Katharina Berndt Rasmussen

Krister Bykvist: *Person-affecting and non-identity*

M.A. Roberts: *What Is the Right Way to Make a Wrong a Right?*

Wlodek Rabinowicz: *Getting Personal – The Intuition of Neutrality Re-interpreted*

Krister Bykvist & Tim Campbell: *Persson's Merely Possible Persons*

Göran Duus-Otterström: *Liability for Emissions without Laws or Political Institutions*

Paul Bowman: *Duties of Corrective Justice and Historical Emissions*

Katie Steele: *The distinct moral importance of acting together*

Gustaf Arrhenius, Mark Budolfson & Dean Spears: *Does Climate Change Policy Depend Importantly on Population Ethics? Deflationary Responses to the Challenges of Population Ethics for Public Policy*

Mark Budolfson & Dean Spears: *Population ethics and the prospects for fertility policy as climate mitigation policy*

Kirsti M. Jylhä, Pontus Strimling & Jens Rydgren: *Climate change denial among radical right-wing supporters*

Malcolm Fairbrother, Gustaf Arrhenius, Krister Bykvist & Tim Campbell: *How Much Do We Value Future Generations? Climate Change, Debt, and Attitudes towards Policies for Improving Future Lives*

Studies on climate ethics and future generations, vol. 3 Working paper series 2021:1–10. Eds. Joe Roussos & Paul Bowman

Dean Spears & H. Orri Stefánsson: *Calibrating Variable-Value Population Ethics*

Joe Roussos: *Awareness Growth and Belief Revision*

Katie Steele: *Why Time Discounting Should Be Exponential: A Reply to Callender*

Nicholas Lawson & Dean Spears: *Population Externalities and Optimal Social Policy*

John Broome: *How Much Harm Does Each of Us Do?*

Tim Campbell: *Offsetting, Denialism, and Risk*

Paul Bowman: *The Relevance of Motivations to Wrongdoing for Contributing to Climate Change*

Hilary Greaves & John Cusbert: *Comparing Existence and Non-Existence*

M.A. Roberts: *Does Climate Change Put Ethics on a Collision Course with Itself?*

Anders Herlitz: *Fixing Person-Based Stakes in Distributive Theory*

Studies on climate ethics and future generations, vol. 4 Working paper series 2021:11–23. Eds. Joe Roussos & Paul Bowman

Gustaf Arrhenius: *Democratic Representation of Future Generations and Population Ethics*

Elizabeth Finneron-Burns: *Global Justice and Future Generations: The Case of Sovereign Wealth Funds*

Paul Bou-Habib & Serena Olsaretti: *Children or Migrants as Public Goods?*

Tim Campbell, Martin Kolk & Julia Mosquera: *Universal Procreation Rights and Future Generations*

H. Orri Stefánsson: *What Is the Point of Offsetting?*

Göran Duus-Otterström: *The Role of Subsistence Emissions in Climate Justice*

Julia Mosquera: *Climate Change, Corrective Justice, and Non-Human Animals*

Göran Duus-Otterström & Edward A. Page: *Defeating Wrongdoing: Why Victims of Unjust Harm Should Take Priority over Victims of Bad Luck*

Henrik Andersson, Eric Brandstedt & Olle Torpman: *Review Article: The Ethics of Population Policies*

Partha Dasgupta & S.J. Beard: *Optimum Population and Environmental Constraints – A Utilitarian Perspective*

M.A. Roberts: *Anonymity and Indefinitely Iterated Addition and Reversal*

Henrik Andersson & Anders Herlitz: *Classifying Comparability Problems in a Way that Matters*

Krister Bykvist & Tim Campbell: *Frick's Defence of the Procreation Asymmetry*

Studies on climate ethics and future generations, vol. 5 Working paper series 2023:1–11. Eds. Tim Campbell & Olle Torpman

Julia Mosquera & Kirsti Jylhä: *How to Feel About Climate Change? An Analysis of the Normativity of Climate Emotions*

John Broome: *How to Value a Person's Life*

Tim Campbell: *DALYs and the Minimally Good Life*

Joe Roussos: *Uncertainty Attitudes as Values in Science*

Karim Jebari & Martin Kolk: *Sex Selection for Daughters: Demographic Consequences of Female-biased Sex Ratios*

Olle Torpman: *Inducement-Based Emissions Accounting*

Stephen M. Gardiner: *The Ethical Risks of an Intergenerational World Climate Bank (as Opposed to a Climate Justice World Bank)*

Katie Steele: *Longtermism and Neutrality about More Lives*

Melinda A. Roberts: *Population, Existence, and Incommensurability*

Patrick Kaczmarek & SJ Beard: *Do We Owe the Past a Future? Reply to Finneron-Burns*

Emil Andersson, Gustaf Arrhenius & Tim Campbell: *Scanlonian Contractualism and Future Generations*

What should we do with regard to climate change given that our choices will not just have an impact on the well-being of future generations, but also determine who and how many people will exist in the future?

There is a very rich scientific literature on different emission pathways and the climatic changes associated with them. There are also a substantial number of analyses of the long-term macroeconomic effects of climate policy. But science cannot say which level of warming we ought to be aiming for or how much consumption we ought to be prepared to sacrifice without an appeal to values and normative principles.

The research program Climate Ethics and Future Generations aims to offer this kind of guidance by bringing together the normative analyses from philosophy, economics, political science, social psychology, and demography. The main goal is to deliver comprehensive and cutting-edge research into ethical questions in the context of climate change policy.

This is the sixth and final volume in our series. It collects eight contributions from researchers in philosophy and political science

Find more information at climateethics.se.

INSTITUTE FOR FUTURES STUDIES

Box 591, SE-101 31
Stockholm, Sweden

Phone:
+46 8 402 12 00

E-mail:
info@iffs.se

Institute for Futures Studies